



Query Match 50.0%; Score 39; DB 2; Length 864;  
Best Local Similarity 40.0%; Pred. No. 76;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ADEEQQALSSQMGF 16  
DB 781 SEEQRQSVQSDQGY 795

RESULT 2  
US-09-022-255-2  
; Sequence 2, Application US/09022255  
; Patent No. 6072033  
; GENERAL INFORMATION:  
; APPLICANT: Yao, Zhengbin  
; APPLICANT: Spriggs, Melanie  
; APPLICANT: Fanslow, William  
; TITLE OF INVENTION: No. 6072033el Receptor That Binds IL-17  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Power Macintosh  
; OPERATING SYSTEM: Apple Operating System 7.5.5  
; SOFTWARE: Microsoft Word for Apple, Version 6.0.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/022,255  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 08/620,694  
; FILING DATE: 21 MARCH 1996  
; APPLICATION NUMBER: USSN 08/538,765  
; FILING DATE: 7 AUGUST 1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 08/410,535  
; FILING DATE: 23 MARCH 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia Anne  
; REGISTRATION NUMBER: 34,695  
; REFERENCE/DOCKET NUMBER: 2617-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)587-0430  
; TELEFAX: (206)

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 864 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-022-255-2

Query Match 50.0%; Score 39; DB 3; Length 864;  
Best Local Similarity 40.0%; Pred. No. 76;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ADEEQQALSSQMGF 16  
DB 781 SEEQRQSVQSDQGY 795

RESULT 3  
US-09-022-696-2  
; Sequence 2, Application US/09022696  
; Patent No. 6072037  
; GENERAL INFORMATION:  
; APPLICANT: Yao, Zhengbin  
; APPLICANT: Spriggs, Melanie  
; APPLICANT: Fanslow, William  
; TITLE OF INVENTION: No. 6072033el Receptor That Binds IL-17  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple PowerMacintosh  
; OPERATING SYSTEM: Apple Operating System 7.5.5

APPLICANT: Yao, Zhengbin  
APPLICANT: Spriggs, Melanie  
APPLICANT: Fanslow, William  
TITLE OF INVENTION: No. 6072037el Receptor That Binds IL-17  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Power Macintosh  
OPERATING SYSTEM: Apple Operating System 7.5.5  
SOFTWARE: Microsoft Word for Apple, Version 6.0.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022,696  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/620,694  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/410,535  
FILING DATE: 23 MARCH 1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia Anne  
REGISTRATION NUMBER: 34,695  
REFERENCE/DOCKET NUMBER: 2617-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 864 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-022-696-2

Query Match 50.0%; Score 39; DB 3; Length 864;  
Best Local Similarity 40.0%; Pred. No. 76;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 2 ADEEQQALSSQMGF 16  
DB 781 SEEQRQSVQSDQGY 795

RESULT 4  
US-08-978-773-2  
; Sequence 2, Application US/08978773  
; Patent No. 6083906  
; GENERAL INFORMATION:  
; APPLICANT: Troutt, Anthony  
; TITLE OF INVENTION: Method of Regulating Nitric Oxide Production  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple PowerMacintosh  
; OPERATING SYSTEM: Apple Operating System 7.5.5

SOFTWARE: Microsoft Word for PowerMacintosh, Version 6.0.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,773  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 60/052,525  
FILING DATE: 27 NOVEMBER 1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia Anne  
REGISTRATION NUMBER: 34,693  
REFERENCE/DOCKET NUMBER: 2623-A  
TELEPHONE: (206)587-0430  
TELEFAX: (206)  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 864 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-978-773-2

Query Match 50.0%; Score 39; DB 3; Length 864;  
Best Local Similarity 40.0%; Pred. No. 76;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 ADEEQQALSSQMGF 16  
Db 781 SEEQRQSVQSDQGY 795

## RESULT 5

US-09-022-253-2  
Sequence 2, Application US/09022253  
Patent No. 6096305  
GENERAL INFORMATION:  
APPLICANT: Yao, Zhengbin  
APPLICANT: Spriggs, Melanie  
APPLICANT: Fanslow, William  
TITLE OF INVENTION: NO. 6096305el Receptor That Binds IL-17  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Power Macintosh  
OPERATING SYSTEM: Apple Operating System 7.5.5  
SOFTWARE: Microsoft Word for Apple, Version 6.0.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022,253  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/620,694  
FILING DATE: 21-MARCH-1996  
APPLICATION NUMBER: USSN 08/538,765  
FILING DATE: 7 AUGUST 1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/410,535  
FILING DATE: 23 MARCH 1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia Anne  
REGISTRATION NUMBER: 34,695

REFERENCE/DOCKET NUMBER: 2617-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 864 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-022-253-2

Query Match 50.0%; Score 39; DB 3; Length 864;  
Best Local Similarity 40.0%; Pred. No. 76;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 ADEEQQALSSQMGF 16  
Db 781 SEEQRQSVQSDQGY 795

## RESULT 6

US-09-022-260-2  
Sequence 2, Application US/09022260  
Patent No. 6100235  
GENERAL INFORMATION:  
APPLICANT: Yao, Zhengbin  
APPLICANT: Spriggs, Melanie  
APPLICANT: Fanslow, William  
TITLE OF INVENTION: NO. 6100235el Receptor That Binds IL-17  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Power Macintosh  
OPERATING SYSTEM: Apple Operating System 7.5.5  
SOFTWARE: Microsoft Word for Apple, Version 6.0.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022,260  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/620,694  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/410,535  
FILING DATE: 23 MARCH 1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia Anne  
REGISTRATION NUMBER: 34,695  
REFERENCE/DOCKET NUMBER: 2617-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 864 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-022-260-2

Query Match 50.0%; Score 39; DB 3; Length 864;





;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/144,121  
;; FILING DATE: 27-OCT-1993  
;; CLASSIFICATION: 424  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Myers, Paul L.  
;; REGISTRATION NUMBER: 35,965  
;; REFERENCE/DOCKET NUMBER: (MGH-0780.0) MGP-021  
;; TELEPHONE: (617) 227-7400  
;; TELEFAX: (617) 227-5941  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 40 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; US-08-144-121-10

Query Match 48.7%; Score 38; DB 1; Length 40;  
Best Local Similarity 53.3%; Pred. No. 3.4;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 ADEEQQALSSOMGF 16  
|:|||||:|  
Db 21 AEGASEQALSAEGF 35

RESULT 10  
US-08-735-893-10  
;; Sequence 10, Application US/08735893  
;; Patent No. 5914317  
;; GENERAL INFORMATION:  
;; APPLICANT: Burgeson, Robert E.  
;; APPLICANT: Wagman, David W.  
;; TITLE OF INVENTION: B1K CHAIN OF LAMININ AND METHODS OF USE  
;; NUMBER OF SEQUENCES: 14  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: LAHIVE & COCKFIELD  
;; STREET: 60 State Street, suite 510  
;; CITY: BOSTON  
;; STATE: Massachusetts  
;; COUNTRY: United States  
;; ZIP: 02109

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/735,893  
;; FILING DATE: 18-OCT-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/144,121  
;; FILING DATE: 27-OCT-1993  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Myers, Paul L.  
;; REGISTRATION NUMBER: 35,965  
;; REFERENCE/DOCKET NUMBER: (MGH-0780.1) MGP-021DV

;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 227-7400  
;; TELEFAX: (617) 227-5941  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 40 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide

;; FRAGMENT TYPE: internal  
;; US-08-735-893-10

Query Match 48.7%; Score 38; DB 2; Length 40;  
Best Local Similarity 53.3%; Pred. No. 3.4;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 ADEEQQALSSOMGF 16  
|:|||||:|  
Db 21 AEGASEQALSAEGF 35

RESULT 11  
US-09-029-213B-25  
;; Sequence 25, Application US/09029213B  
;; Patent No. 6180098  
;; GENERAL INFORMATION:  
;; APPLICANT: CHRISTIAN, Peter D.  
;; TITLE OF INVENTION: RECOMBINANT HELICOVERPA BACULOVIRUSES  
;; TITLE OF INVENTION: EXPRESSING HETEROLOGOUS DNA  
;; NUMBER OF SEQUENCES: 27  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: McDermott, Will & Emery  
;; STREET: 600 13th Street, NW  
;; CITY: Washington  
;; STATE: District of Columbia  
;; COUNTRY: USA  
;; ZIP: 20005

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/029,213B  
;; FILING DATE: 31-AUG-1998  
;; CLASSIFICATION: 424

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Joseph Hyosuk Kim  
;; REGISTRATION NUMBER: 41,425  
;; REFERENCE/DOCKET NUMBER: 50179-048  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 202-756-8000  
;; TELEFAX: 202-756-8087  
;; INFORMATION FOR SEQ ID NO: 25:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 323 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-029-213B-25

Query Match 48.7%; Score 38; DB 4; Length 323;  
Best Local Similarity 58.3%; Pred. No. 37;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 DEEQQALSSQM 14  
|:|||||:|  
Db 168 DDEQQQASSEL 179

RESULT 12  
US-08-560-398-2  
;; Sequence 2, Application US/08560398  
;; Patent No. 5907082  
;; GENERAL INFORMATION:  
;; APPLICANT: O'Neill, Sharmen  
;; APPLICANT: Nadeau, Jeanette  
;; TITLE OF INVENTION: Ovule-Specific Gene Expression  
;; NUMBER OF SEQUENCES: 13  
;; CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/560,398  
FILING DATE: 17-NOV-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-063300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 768 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-560-398-2

Query Match 48.7%; Score 38; DB 2; Length 768;  
Best Local Similarity 46.2%; Pred. No. 99;  
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 3 DEEQQALSSQMG 15  
|::|::|::|::|  
Db 115 DDKQKALSRELG 127

RESULT 13  
US-08-620-694A-10  
Sequence 10, Application US/08620694A  
Patent No. 5869286  
GENERAL INFORMATION:  
APPLICANT: Yao, Zhengbin  
APPLICANT: Spriggs, Melanie  
APPLICANT: Fanslow, William  
TITLE OF INVENTION: No. 5869286el Receptor That Binds IL-17  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Power Macintosh  
OPERATING SYSTEM: Apple Operating System 7.5.5  
SOFTWARE: Microsoft Word for Apple, Version 6.0.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/620,694A  
FILING DATE: 21 MARCH 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/538,765  
FILING DATE: 7 AUGUST 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/410,535  
FILING DATE: 23 MARCH 1995

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia Anne  
REGISTRATION NUMBER: 34,695  
REFERENCE/DOCKET NUMBER: 2617-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 866 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-620-694A-10

Query Match 48.7%; Score 38; DB 2; Length 866;  
Best Local Similarity 42.9%; Pred. No. 11e+02;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 DEEQQALSSQMGF 16  
|::|::|::|::|  
Db 783 EEEQROSVQSDQGY 796

RESULT 14  
US-09-022-255-10  
Sequence 10, Application US/09022255  
Patent No. 6072033  
GENERAL INFORMATION:  
APPLICANT: Yao, Zhengbin  
APPLICANT: Spriggs, Melanie  
APPLICANT: Fanslow, William  
TITLE OF INVENTION: No. 6072033el Receptor That Binds IL-17  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Power Macintosh  
OPERATING SYSTEM: Apple Operating System 7.5.5  
SOFTWARE: Microsoft Word for Apple, Version 6.0.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022,255  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/620,694  
FILING DATE: 21 MARCH 1996  
APPLICATION NUMBER: USSN 08/538,765  
FILING DATE: 7 AUGUST 1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/410,535  
FILING DATE: 23 MARCH 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia Anne  
REGISTRATION NUMBER: 34,695  
REFERENCE/DOCKET NUMBER: 2617-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 866 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-022-255-10

Job time: 316 sec

Query Match 48.7%; Score 38; DB 3; Length 866;  
Best Local Similarity 42.9%; Pred. No. 1.1e+02;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 DEEQOALSSQMGF 16  
:||||:| |:  
Db 783 EEEQSQVSDQY 796

RESULT 15

US-09-022-696-10  
; Sequence 10, Application US/09022696  
; Patent No. 6072037  
; GENERAL INFORMATION:  
; APPLICANT: Yao, Zhengbin  
; APPLICANT: Spriggs, Melanie  
; APPLICANT: Fanslow, William  
; TITLE OF INVENTION: No. 6072037el Receptor That Binds IL-17  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Power Macintosh  
; OPERATING SYSTEM: Apple Operating System 7.5.5  
; SOFTWARE: Microsoft Word for Apple, Version 6.0.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/022,696  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION NUMBER: 08/620,694  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 08/410,535  
; FILING DATE: 23 MARCH 1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia Anne  
; REGISTRATION NUMBER: 34,695  
; REFERENCE/DOCKET NUMBER: 2617-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)587-0430  
; TELEFAX: (206)  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 866 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-022-696-10

Query Match 48.7%; Score 38; DB 3; Length 866;  
Best Local Similarity 42.9%; Pred. No. 1.1e+02;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 DEEQOALSSQMGF 16  
:||||:| |:  
Db 783 EEEQSQVSDQY 796

Search completed: July 5, 2001, 11:47:02

7

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:48:37 ; Search time 79.63 seconds  
(without alignments)  
15.306 Million cell updates/sec

Title: US-09-462-480-13  
Perfect score: 78  
Sequence: 1 RADEEQQALSSQMGF 16  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	100	2 H70802	hypothetical prote
2	49	1263	2	T15496	hypothetical prote
3	43	55.1	745	2 D96829	homeobox protein (
4	43	55.1	747	2 S71478	homeotic protein A
5	43	55.1	826	1 A31822	villin - chicken
6	42	53.8	100	2 T10032	hypothetical prote
7	42	53.8	224	2 T35682	two-component regu
8	42	53.8	902	2 G83635	probable ClpA/B-ty
9	41	52.6	664	2 B53610	ntpl protein - Ent
10	41	52.6	856	2 T34491	hypothetical prote
11	40.5	51.9	258	2 A83169	O-acetylserine syn
12	40	51.3	687	2 S19680	protein-glutamine
13	40	51.3	911	2 S70958	otnA protein - Vib
14	39	50.0	355	2 F86328	hypothetical prote
15	39	50.0	398	2 S44028	actin-related prot
16	39	50.0	682	2 S71476	hypothetical prote
17	39	50.0	827	2 T20492	homeotic protein H
18	39	50.0	838	2 T04449	hypothetical prote
19	39	50.0	1286	2 T16507	hypothetical prote
20	38.5	49.4	83	2 T08238	gas-vesicle operon
21	38.5	49.4	83	2 JQ1124	tail fiber assembl
22	38	48.7	175	2 C42291	hypothetical prote
23	38	48.7	175	2 B42463	hypothetical prote
24	38	48.7	198	2 T13215	hypothetical prote
25	38	48.7	381	2 T34333	hypothetical prote
26	38	48.7	397	2 F82650	hypothetical prote
27	38	48.7	529	2 S76831	hypothetical prote
28	38	48.7	564	2 H83397	probable two-compo
29	38	48.7	571	2 S48384	DNA43 protein - ye

30	38	48.7	601	2 D83583	probable acyl-CoA
31	38	48.7	768	2 S71477	homeotic protein,
32	38	48.7	1170	2 A53612	laminin B1k chain
33	38	48.7	1805	2 A34736	nestin - rat
34	37	47.4	94	2 A25833	carboxypeptidase A
35	37	47.4	149	2 S74964	hypothetical prote
36	37	47.4	197	2 F96799	similar to 'MADS b
37	37	47.4	281	2 A37349	chorion protein s3
38	37	47.4	283	2 A60364	tropomyosin - migr
39	37	47.4	285	2 C25242	tropomyosin, exon
40	37	47.4	285	2 A25561	tropomyosin II, mu
41	37	47.4	286	2 B25242	tropomyosin, exon
42	37	47.4	286	2 A25242	tropomyosin, exon
43	37	47.4	359	2 F82541	dihydroorotate deh
44	37	47.4	386	2 T05691	multiubiquitin cha
45	37	47.4	405	2 T45615	hypothetical prote

## ALIGNMENTS

### RESULT 1

H70802  
hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70802  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A: Reference number: A70500; MUID: 98295987  
A: Accession: H70802  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-100 <COL>  
A: Cross-references: GB:AL022120; GB:AL123456; NID:g3761558; PIDN:CAA17966.1; PID:g296  
A: Experimental source: strain H37RV  
A: Genetics:  
A: Gene: RV3874

Query Match 100.0%; Score 78; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 2e-06; 0; Indels 0; Gaps  
Matches 16; Conservative 0; Mismatches 0

QY 1 RADEEQQALSSQMGF 16  
|||||  
DB 85 RADEEQQALSSQMGF 100

### RESULT 2

T15496  
hypothetical protein C14F5.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15496  
R: Minx, P.  
submitted to the EMBL Data Library, June 1995  
A: Description: The sequence of C. elegans cosmid C14F5.  
A: Reference number: Z18361  
A: Accession: T15496  
A: Status: preliminary; translated from GB/EMBL/DBJ  
A: Molecule type: DNA  
A: Residues: 1-1263 <MIN>  
A: Cross-references: EMBL:U29082; NID:g861384; PID:g861386; PIDN:AAA68402.1; CESP:C14F  
A: Experimental source: strain Bristol N2  
C: Genetics:  
A: Gene: CESP:C14F5.3  
A: Introns: 21/3; 34/2; 740/2; 852/2; 881/2; 902/3; 970/3; 1005/2; 1041/3; 1093/3; 112

```

A:Accession: AJ3500
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 3-647,'RSRRYKHDTAGAR',661,'T' <RER>
A:Cross-references: GB:I32873
A:Note: this sequence has been revised in reference S71449
C:Genetics:
A:Gene: GL2
A:Introns: 26/3; 118/2; 157/3; 336/3; 371/3; 502/1; 592/3; 621/3
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:102-158/Domain: homeobox homology <HOX>

Query Match          55.1%; Score 43; DB 2: Length 747;
Best Local Similarity 61.5%; Pred. No. 20;
Matches      8; Conservative      3; Mismatches      2; Indels      0;

QY      3 DEQQQALSSQMG 15
      ||:|: || |
Db      127 DEKQRLSKQLG 139

RESULT      5
A31822
villin - chicken
N:Alternate names: villin 1
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-S
C:Accession: A31822; A25703; A03082; A42816
R:Bazari, W.L.; Matsudaira, P.; Wallek, M.; Smeal, T.; Jakes, R.; Ahn
Proc. Natl. Acad. Sci. U.S.A. 85, 4986-4990, 1988
A:Title: Villin sequence and peptide map identify six homologous doma
A:Reference number: A31822; MUID:88276884
A:Accession: A31822

```

A; Residues: 1-826 <BAZ>  
A; Cross-references: GB:J03781; NID:Q212862; PIDN:AAA49133.1; PID:Q212862  
R; Pringault, E.; Arpin, M.; Garcia, A.; Finidori, J.; Louvard, D.  
EMBO J. 5, 3119-3124, 1986  
A; Title: A human villin cDNA clone to investigate the differentiation of the villi  
A; Reference number: A91057; MUID:87133467  
A; Accession: A25703  
A; Molecule type: mRNA  
A; Residues: 751-826 <PRI>  
R; Glennay Jr., J.R.; Geisler, N.; Kaulfus, P.; Weber, K.  
J. Biol. Chem. 256, 8156-8161, 1981  
A; Title: Demonstration of at least two different actin-binding sites in villin  
A; Reference number: A92333; MUID:81264203  
A; Accession: A03082  
A; Molecule type: protein  
A; Residues: 751-826 <GLE>  
R; Herterberg, L.K.; Weber, K.  
J. Biol. Chem. 258, 365-369, 1983  
A; Title: Demonstration of three distinct calcium-binding sites in villin  
A; Reference number: A92418; MUID:83082892  
A; Contents: annotation  
A; Note: Gly-32, Asp-34, Ser-36, Lys-38, Asn-40, and Ser-43 may be involved in calcium binding  
R; de Arruda, M.V.; Bazari, H.; Wallek, M.; Matsudaira, P.  
J. Biol. Chem. 267, 13079-13085, 1992  
A; Title: An actin footprint on villin. Single site substitutions in a villin cDNA  
A; Reference number: A42816; MUID:92317005  
A; Accession: A42816  
A; Status: preliminary  
A; Molecule type: protein  
A; Residues: 1-11; 128-132 <DEA>  
C; Comment: Villin is abundant in the microvilli of intestinal epithelial cells  
C; Function:  
C; Description: calcium-regulated actin-binding; in vitro, bundles actin  
C; Superfamily: villin; gelsolin repeat homology; villin headpiece homology  
C; Keywords: actin binding; calcium binding; duplication  
F; 2-826/Product: villin #status experimental <MAT>  
F; 17-350/Domain: gelsolin repeat homology <GEL1>

F:17-49,398-430/Region: repeats a, a'  
F:62-78,174-190,295-311,443-459,554-570,658-674/Region: repeats b, c, d, b', c', d'  
F:398-715/Domain: gelsolin repeat homology <GEL2>  
F:763-826/Domain: villin headpiece homology <VHH>

Query Match 55.1%; Score 43; DB 1; Length 826;  
Best Local Similarity 61.5%; Pred. No. 22;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 EEOQQALSSOMGF 16  
||:||||:|  
Db 306 EEKQAMSRALGF 318

RESULT 6

T10032

hypothetical protein MCB628.13c - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000

C:Accession: T10032

R:Bigliardi, K.; Honore, N.; Woods, S.A.; Caudron, B.; Cole, S.T.

Mol. Microbiol. 7, 197-206, 1993

A:Title: Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae

A:Reference number: Z16917; MUID:93188700

A:Accession: T10032

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-100 <BIG>

A:Cross-references: EMBL:Y14967; NID:g2370268; PIDN:CAA75210.1; PID:g2370280

Query Match 53.8%; Score 42; DB 2; Length 100;  
Best Local Similarity 50.0%; Pred. No. 3.1;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 RADEEQQALSSOMGF 16  
:|:|:|:|  
Db 85 KTDDEANQLLSKMF 100

RESULT 7

T35682

two-component regulator - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 31-Jan-2000

C:Accession: T35682

R:Harris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1998

A:Reference number: Z21587

A:Accession: T35682

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-224 <HAR>

A:Cross-references: EMBL:AL031031; PIDN:CAA19850.1; GSPDB:GN00070; SCOEDB:SC7C7.04

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC7C7.04

C:Superfamily: ompR protein; response regulator homology

Query Match 53.8%; Score 42; DB 2; Length 224;  
Best Local Similarity 64.3%; Pred. No. 7.7;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ADEEQQALSSOMG 15  
|:|:|:|:|  
Db 170 AVEEQAEALSKQLG 183

RESULT 8

G83635

probable ClpA/B-type chaperone PA0090 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83635  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: G83635  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-902 <STO>  
A:Cross-references: GB:AE004447; GB:AE004091; NID:g9945902; PIDN:AAG03480.1; GSPDB:GN00070  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0090  
C:Superfamily: endopeptidase Clp ATP-binding chain

Query Match 53.8%; Score 42; DB 2; Length 902;  
Best Local Similarity 57.1%; Pred. No. 36;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RADEEQQALSSOM 14  
|:|:|:|:|  
Db 532 RLDEEQQALHARL 545

RESULT 9

B53610

ntpI protein - Enterococcus hirae

C:Species: Enterococcus hirae

C:Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 01-Dec-2000

C:Accession: B53610; B54392

R:Takase, K.; Kakinuma, S.; Yamato, I.; Konishi, K.; Igarashi, K.; Kakinuma, Y.

J. Biol. Chem. 269, 11037-11044, 1994

A:Title: Sequencing and characterization of the ntp gene cluster for vacuolar-type Na<sup>+</sup> ATPase of Enterococcus hirae

A:Reference number: A53610; MUID:94209269

A:Accession: B53610

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-664 <TAB>

A:Cross-references: GB:D17462; NID:g487271; PIDN:BAA04270.1; PID:d1004790; PID:g48727

R:Solioz, M.; Davies, K.

J. Biol. Chem. 269, 9453-9459, 1994

A:Title: Operon of vacuolar-type Na<sup>+</sup> ATPase of Enterococcus hirae.

A:Reference number: A54392; MUID:94193617

A:Accession: B54392

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-500, 'G', 502-514, 'I', 516-664 <SOL>

A:Cross-references: GB:X76913; NID:g472916; PIDN:CAA54236.1; PID:g472918

C:Genetics:

A:Gene: ntpM

Query Match 52.6%; Score 41; DB 2; Length 664;  
Best Local Similarity 43.8%; Pred. No. 38;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 RADEEQQALSSOMGF 16  
|:|:|:|:|  
Db 205 RADESMEIASRYGF 220

RESULT 10

T34491

hypothetical protein ZK1248.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T34491

R:Latreille, P.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of *C. elegans* cosmid ZK1248.

A:Reference number: Z21534

A:Accession: T34491

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-856 <LAT>

A:Cross-references: EMBL:U29244; PIDN:AACT1091.1; GSPDB:GN00020; CESP:ZK1248.10

A:Experimental source: strain Bristol N2; clone ZK1248

C:Genetics:

A:Gene: CESP:ZK1248.10

A:Map position: 2

A:Introns: 26/3; 86/2; 224/3; 271/3; 317/2; 445/2; 527/1; 682/3; 803/1; 850/3

Query Match 52.6%; Score 41; DB 2; Length 856;

Best Local Similarity 50.0%; Pred. No. 51;

Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RADEEQQALSSQMGF 16

Db 542 RTPESQNLTDQLGF 557

RESULT 11

A83169

O-acetylserine synthase PA3816 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 16-Feb-2001

C:Accession: A83169

R:Stover, C.K.; Pham, X.Q.; Ervin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: A83169

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <STO>

A:Cross-references: GB:AE004799; GB:AE004091; NID:g9949981; PIDN:AAG07203.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: cySE; PA3816

C:Superfamily: *Bacillus* serine acetyltransferase; serine acetyltransferase homology

Query Match

Best Local Similarity 47.4%; Score 40.5; DB 2; Length 258;

Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

Qy 1 RADEEQQALSSQMGF 16

Db 170 REDSEQAKQMAEKLGF 188

RESULT 12

S19680

protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) - bovine

N:Alternate names: 25K retinol-induced protein

C:Species: *Bos primigenius taurus* (cattle)

C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 22-Oct-1999

C:Accession: S19680; A34446

R:Nakanishi, K.; Nara, K.; Hagiwara, B.; Aoyama, Y.; Ueno, H.; Hirose, S.

Eur. J. Biochem. 202, 15-21, 1991

A:Title: Cloning and sequence analysis of cDNA clones for bovine aortic-endothelial-cell

A:Reference number: S19680; MUID:52037637

A:Accession: S19680

A:Molecule type: mRNA

A:Residues: 1-687 <NAK>

A:Cross-references: EMBL:X60686; NID:g817; PIDN:CAA43097.1; PID:g818

R:Nara, K.; Nakanishi, K.; Hagiwara, H.; Wakita, K.; Kojima, S.; Hirose, S.

J. Biol. Chem. 264, 19308-19312, 1989

A:Title: Retinol-induced morphological changes of cultured bovine endothelial cells a

A:Reference number: A34446; MUID:90037069

A:Accession: A34446

A:Status: preliminary

A:Molecule type: protein

A:Residues: 79-95;157-166;242-251 <NAR>

C:Superfamily: protein-glutamine gamma-glutamyltransferase

C:Keywords: aminoacyltransferase

F:2/77/Active site: Cys #status predicted

Query Match 51.3%; Score 40; DB 2; Length 687;

Best Local Similarity 46.7%; Pred. No. 59;

Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ADEEQQALSSQMGF 16

Db 152 SDEERQYVLTKQGF 166

RESULT 13

S70958

otna protein - *Vibrio cholerae*

C:Species: *Vibrio cholerae*

C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 08-Oct-1999

C:Accession: S70958

R:Bik, E.M.; Bunschoten, A.E.; Willems, R.J.L.; Chang, A.C.Y.; Mooi, F.R.

Mol. Microbiol. 20, 799-811, 1996

A:Title: Genetic organization and functional analysis of the otn DNA essential for ce

A:Reference number: S70952; MUID:96386047

A:Accession: S70958

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-911 <BIK>

A:Cross-references: EMBL:X90547; NID:g1469276; PIDN:CAA62140.1; PID:e194955; PID:g110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

C:Genetics:

A:Gene: otnA

Query Match

Best Local Similarity 72.7%; Score 40; DB 2; Length 911;

Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 6 QQQALSSQMGF 16

Db 41 QQQALASKYGF 51

RESULT 14

F86328

hypothetical protein AAF98405.1 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: F86328

R:Neologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719

A:Accession: F86328

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-355 <STO>

A:Cross-references: GB:AE005172; NID:g9795587; PIDN:AAF98405.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1



Query Match 50.0%; Score 39; DB 2; Length 355;  
 Best Local Similarity 57.1%; Pred. No. 42;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 DEEQQALSSQMGF 16  
 I:| | | | |  
 Db 100 DDEQGLFVSNNMF 113

RESULT 15

S44028  
 actin-related protein - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 21-Jul-2000  
 C:Accession: S44028  
 R:Frankel, S.; Heintzelman, M.B.; Artavanis-Tsakonas, S.; Mooseker, M.S.  
 J. Mol. Biol. 235, 1351-1356, 1994  
 A:Title: Identification of a divergent actin-related protein in Drosophila.  
 A:Reference number: S44028; MUID:94141930  
 A:Accession: S44028  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-398 <FRA>  
 A:Cross-references: EMBL:L25314; NID:g408943; PIDN:AAAI7685.1; PID:g408944  
 C:Genetics:  
 A:Gene: FlyBase:Actrl3E  
 A:Cross-references: FlyBase:FBgn0011741  
 C:Superfamily: actin

Query Match 50.0%; Score 39; DB 2; Length 398;  
 Best Local Similarity 72.7%; Pred. No. 48;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RADEEQQALS 11  
 I:| | | | |  
 Db 264 REDEEQQMV 274

Search completed: July 5, 2001, 11:48:38  
 Job time: 377 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:51:46 ; Search time 41.8 Seconds  
(without alignments)  
13.112 Million cell updates/sec

Title: US-09-462-480-13

Perfect score: 78

Sequence: 1 RADEEQQALSSQMGF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	43	55.1	745	1	HGL2_ARATH
2	43	55.1	826	1	VILLI_CHICK
3	41	52.6	664	1	NTPI_ENTHR
4	40	51.3	687	1	TGLC_BOVIN
5	39	50.0	398	1	ACTU_DROME
6	38.5	49.4	83	1	GVGI_HALNI
7	38	48.7	175	1	TFA_BPMU
8	38	48.7	175	1	TFA_BPP2
9	38	48.7	571	1	DN43_YEAST
10	38	48.7	1172	1	LMB3_HUMAN
11	38	48.7	1805	1	NEST_RAT
12	37	47.4	94	1	CBPA_PIG
13	37	47.4	283	1	TPNM_LOCFI
14	37	47.4	285	1	TPM1_DROME
15	37	47.4	386	1	PSD4_ARATH
16	37	47.4	419	1	CBPA_BOVIN
17	37	47.4	438	1	TPM1_DROME
18	37	47.4	504	1	TPM4_DROME
19	37	47.4	518	1	SELB_DROME
20	37	47.4	634	1	SELB_WOOTH
21	37	47.4	689	1	TGLC_CAVCU
22	37	47.4	1021	1	YL88_CAEEL
23	36.5	46.8	838	1	AXN2_RAT
24	36.5	46.8	840	1	AXN2_MOUSE
25	36	46.2	212	1	RB15_RAT
26	36	46.2	274	1	P30_MYCPN
27	36	46.2	361	1	RIBD_BACSU
28	36	46.2	433	1	TIG_VIBCH
29	36	46.2	440	1	Y788_BORBU
30	36	46.2	468	1	YOC1_CAEEL
31	36	46.2	634	1	SELB_DESBA
32	36	46.2	687	1	TGLC_HUMAN
33	36	46.2	702	1	EPG_BUCAI

34	36	46.2	1168	1	LMB3_MOUSE	Q61087 mus musculus
35	36	46.2	1377	1	RHSA_ECOLI	P16916 escherichia
36	36	46.2	1397	1	RHSC_ECOLI	P16918 escherichia
37	36	46.2	1411	1	RHSB_ECOLI	P16917 escherichia
38	35.5	45.5	543	1	VIBE_VIBCH	O07899 vibrio chol
39	35	44.9	282	1	YIAJ_ECOLI	P37671 escherichia
40	35	44.9	302	1	YCAP_CAEEL	Q22915 caenorhabdi
41	35	44.9	313	1	GDA7_WHEAT	P04727 triticum ae
42	35	44.9	319	1	GDA5_WHEAT	P04725 triticum ae
43	35	44.9	325	1	RECA_PROVU	P26346 proteus vul
44	35	44.9	342	1	RECA_ERWCA	P26344 erwinia car
45	35	44.9	346	1	BI0B_ERWHE	Q47862 erwinia her

## ALIGNMENTS

RESULT 1  
HGL2\_ARATH  
ID HGL2\_ARATH STANDARD; PRT; 745 AA.  
AC P46607: Q39018:  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HOMEOBOX PROTEIN GLABRA2 (HOMEOBOX-LEUCINE ZIPPER PROTEIN ATHB-10)  
DE (HD-ZIP PROTEIN ATHB-10).  
GN GL2 OR FL9K16.20.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. WASSILEWSKIIA; TISSUE=Seedling;  
RX MEDLINE=95011550; PubMed=7926739;  
RA Rerie W.G., Feldmann K.A., Marks M.D.;  
RT "The GLABRA2 gene encodes a homeo domain protein required for normal  
RT trichome development in Arabidopsis.";  
RL Genes Dev. 8:1388-1399(1994).  
RN [2]  
RP REVISIONS.  
RA Marks M.D.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=96407838; PubMed=8811855;  
RA di Cristina M., Sessa G., Dolan L., Linstead P., Balma S., Ruberti I.,  
RA Morelli G.;  
RT "The Arabidopsis Athb-10 (GLABRA2) is an HD-Zip protein required for  
RT regulation of root hair development.";  
RL Plant J. 10:393-402(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=96407838; PubMed=8811855;  
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,  
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,  
RA Barnstead M.E., Bowman C.L., White O., Niernan W.C., Fraser C.M.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: REQUIRED FOR CORRECT MORPHOLOGICAL DEVELOPMENT AND  
CC MATURATION OF TRICHOMES AS WELL AS FOR NORMAL DEVELOPMENT OF SEED  
CC COAT MUCILAGE.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING TRICHOMES.  
CC -!- SIMILARITY: BELONGS TO THE HD-ZIP FAMILY OF HOMEOBOX PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L32873; AAC80260.1; -
CC DR EMBL; Z54356; CAA91183.1; ALT_INIT.
CC DR EMBL; AC01117; AAF09047.1; -
CC DR HSP; P02836; 2HDD.
CC DR InterPro; IPR001356; -
CC DR InterPro; IPR002913; -
CC DR Pfam; PF01852; START; 1.
CC DR PROSITE; PS00027; HOMEBOX_1; 1.
CC DR PROSITE; PS00046; homeobox; 1.
CC DR PROSITE; PS00071; HOMEBOX_2; 1.
CC DR HOMEBOX; DNA-binding; Nuclear protein.
KW DOMAIN 39 60 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 70 87 ASP/GLU-RICH (ACIDIC).
FT DNAS_BIND 99 158 HOMEBOX
FT SEQUENCE 745 AA; 82988 MW; E4AD2DAAC155DF3 CRC64;

Query Match 55.1%; Score 43; DB 1; Length 745;
Best Local Similarity 61.5%; Pred. No. 12;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 DEEQOQALSSQMG 15
Db 125 DEKQOQLSKQLG 137

RESULT 2
VILLI_CHICK STANDARD; PRT; 826 AA.
AC P02640;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE VILLIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archozoa; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88276884; PubMed=2839826;
RA Bazari W.L., Matsudaira P., Wallik M., Smeal T., Jakes R., Ahmed Y.;
RT "Villin sequence and peptide map identify six homologous domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:4986-4990(1988).
RN [2]
RP SEQUENCE OF 751-826.
RX MEDLINE=81264203; PubMed=6790532;
RA Glenney J.R. Jr., Geisler N., Kaulfus P., Weber K.;
RT "Demonstration of at least two different actin-binding sites in villin, a calcium-regulated modulator of F-actin organization.";
RL J. Biol. Chem. 256:8156-8161(1981).
RN [3]
RP CALCIUM-BINDING SITES.
RX MEDLINE=83082892; PubMed=6848508;
RA Hesterberg L.K., Weber K.;
RT "Demonstration of three distinct calcium-binding sites in villin, a modulator of actin assembly.";
RL J. Biol. Chem. 258:365-369(1983).
RN [4]
RP STRUCTURE BY NMR OF 1-127.
RX MEDLINE=94191534; PubMed=8142900;
RA Markus M.A., Nakayama T., Matsudaira P., Wagner G.;
RT "Solution structure of villin 147, a domain conserved among actin-severing proteins.";
RL Protein Sci. 3:70-81(1994).
RN [5]
RP STRUCTURE BY NMR OF 1-127.
RX MEDLINE=97337440; PubMed=9194180;
RA Markus M.A., Matsudaira P., Wagner G.;
RT "Refined structure of villin 147 and a detailed comparison with other actin-severing domains.";
RN Protein Sci. 6:1197-1209(1997).
RN [6]
RP STRUCTURE BY NMR OF 792-826.
RX MEDLINE=97307248; PubMed=9164455;
RA McKnight C.J., Matsudaira P.T., Kim P.S.;
RT "NMR structure of the 35-residue villin headpiece subdomain.";
RL Nat. Struct. Biol. 4:180-184(1997).
CC -1- FUNCTION: VILLIN IS A CA(2+)-REGULATED ACTIN-BINDING PROTEIN. ITS CAPPING FUNCTION IS LOCALIZED TO DOMAIN I; ITS SEVERING ACTION IS A COOPERATION BETWEEN AN F-ACTIN BINDING SITE OF DOMAIN II AND THE CAPPING ACTIVITY OF DOMAIN I.
CC -1- SUBUNIT: MONOMER.
CC -1- TISSUE SPECIFICITY: COMPONENT OF BRUSH BORDER MICROVILLI.
CC -1- SIMILARITY: BELONGS TO THE VILLIN/GELSOLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J03781; AAA49133.1; -
CC DR PIR; A03082; A03082.
CC DR PIR; A31822; A31822.
CC DR PDB; 2VIK; 01-APR-97.
CC DR PDB; 2VII; 01-APR-97.
CC DR PDB; 1VII; 12-AUG-97.
CC DR InterPro; IPR001974; -
CC DR Pfam; PF00626; Gelsolin; 6.
CC DR PRINTS; PR00597; GELSOLIN.
KW Calcium; Actin-binding; Capping protein; Repeat; 3D-structure.
FT DOMAIN 1 734 CORE.
FT DOMAIN 735 826 HEADPIECE.
FT REPEAT 7 369 1ST HALF.
FT DOMAIN 370 386 HINGE REGION.
FT REPEAT 387 734 2ND HALF.
FT DOMAIN 7 126 1A.
FT REPEAT 40 48 MOTIF B.
FT REPEAT 62 78 MOTIF A.
FT REPEAT 100 111 MOTIF C.
FT DOMAIN 127 249 1B.
FT REPEAT 160 168 MOTIF B.
FT REPEAT 174 190 MOTIF A.
FT REPEAT 216 226 MOTIF C.
FT DOMAIN 250 369 1C.
FT REPEAT 280 288 MOTIF B.
FT REPEAT 295 311 MOTIF A.
FT REPEAT 335 346 MOTIF C.
FT DOMAIN 387 507 2A.
FT REPEAT 421 429 MOTIF B.
FT REPEAT 443 459 MOTIF A.
FT REPEAT 481 492 MOTIF C.
FT DOMAIN 508 614 2B.
FT REPEAT 540 548 MOTIF B.
FT REPEAT 554 570 MOTIF A.
FT REPEAT 587 598 MOTIF C.
FT DOMAIN 615 734 2C.
FT REPEAT 644 652 MOTIF B.
FT REPEAT 658 673 MOTIF A.
FT REPEAT 700 712 MOTIF C.
FT SITE 820 823 ABSOLUTELY REQUIRED FOR ACTIVITY.
FT DOMAIN 129 137 CRUCIAL FOR BINDING AN ACTIN FILAMENT.
FT CA_BIND 782 793 PROBABLE.
FT DOMAIN 112 119 POLYPHOSPHONOSITIDE BINDING (BY SIMILARITY).
FT DOMAIN 138 146 POLYPHOSPHONOSITIDE BINDING (BY SIMILARITY).
FT STRAND 21 23
FT TURN 30 32
FT TURN 34 36
```

FT STRAND 39 41  
 FT STRAND 44 44  
 FT STRAND 47 53  
 FT STRAND 58 64  
 FT STRAND 68 68  
 FT STRAND 72 88  
 FT TURN 89 90  
 FT STRAND 95 97  
 FT STRAND 104 110  
 FT TURN 112 113  
 FT STRAND 116 118  
 FT TURN 124 125  
 SQ SEQUENCE 826 AA; 92479 MW; 6A8898F7DF947389 CRC64;

Query Match 55.1%; Score 43; DB 1; Length 826;  
 Best Local Similarity 61.5%; Pred. No. 14;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 4 EEOQQALSSQMGF 16  
 II:III:I :II  
 DB 306 EEKQAMSRALGF 318

RESULT 3  
 ID NTPI\_ENTHR STANDARD; PRT; 664 AA.  
 AC P43439;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE V-TYPE SODIUM ATP SYNTHASE SUBUNIT I (EC 3.6.1.34) (NA(+)-  
 DE TRANSLLOCATING ATPASE SUBUNIT I).  
 GN NTPI OR NTPM.  
 OS Enterococcus hirae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;  
 OC Enterococcus.  
 OX NCBI\_TaxID=1354;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 9790;  
 RX MEDLINE=94209269; PubMed=8157629;  
 RA Takase K., Kakinuma S., Yamato I., Konishi K., Igarashi K.,  
 RA Kakinuma Y.;  
 RT "Sequencing and characterization of the ntp gene cluster for  
 RT vacuolar-type Na(+)-translocating ATPase of Enterococcus hirae.";  
 RL J. Biol. Chem. 269:11037-11044(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 9790;  
 RX MEDLINE=94193617; PubMed=8144530;  
 RA Solioz M., Davies K.;  
 RT "Operon of vacuolar-type Na(+)-ATPase of Enterococcus hirae.";  
 RL J. Biol. Chem. 269:9453-9459(1994).  
 CC -1- FUNCTION: INVOLVED IN ATP-DRIVEN SODIUM EXTRUSION.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D17462; BAA04270.1;  
 CC EMBL; X76913; CAA54236.1;  
 CC InterPro; IPR002490;  
 CC Pfam; PF01496; V\_ATPase\_sub\_a; 1.  
 KW Hydrolase; Hydrogen ion transport; Transmembrane.  
 FT TRANSMEM 373 393 POTENTIAL.  
 FT TRANSMEM 460 480 POTENTIAL.  
 FT TRANSMEM 496 516 POTENTIAL.

FT TRANSMEM 522 542 POTENTIAL.  
 FT TRANSMEM 574 594 POTENTIAL.  
 FT TRANSMEM 601 621 POTENTIAL.  
 FT CONFLICT 501 515 W -> G (IN REF. 2).  
 FT CONFLICT 515 515 T -> I (IN REF. 2).  
 SQ SEQUENCE 664 AA; 75620 MW; 08CC146D6CB89AF5 CRC64;

Query Match 52.6%; Score 41; DB 1; Length 664;  
 Best Local Similarity 43.8%; Pred. No. 23;  
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 RADEEQQALSSQMGF 16  
 IIII: : :II:  
 DB 205 RADESMEEIASRYGF 220

RESULT 4  
 ID TGLC\_BOVIN STANDARD; PRT; 687 AA.  
 AC P51176;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE PROTEIN-GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE (EC 2.3.2.13) (TISSUE  
 DE TRANSGLUTAMINASE) (TGASE C) (TGC).  
 GN TGM2.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 79-95; 157-166; 242-251 & 581-587.  
 RC TISSUE=artery;  
 RX MEDLINE=92037637; PubMed=1682150;  
 RA Nakanishi K., Nara K., Hagiwara H., Aoyama Y., Ueno H., Hirose S.;  
 RT "Cloning and sequence analysis of cDNA clones for bovine aortic-  
 RT endothelial-cell transglutaminase.";  
 RL Eur. J. Biochem. 202:15-21(1991).  
 CC -1- FUNCTION: CATALYZES THE CROSS-LINKING OF PROTEINS AND THE  
 CC CONJUGATION OF POLYAMINES TO PROTEINS.  
 CC -1- CATALYTIC ACTIVITY: PROTEIN GLUTAMINE + ALKYLAMINE -> PROTEIN  
 CC N(5)-ALKYLGLUTAMINE + NH(3).  
 CC -1- COFACTOR: CALCIUM IS REQUIRED FOR TRANSGLUTAMINASE ACTIVITY.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS ARE DETECTED IN THE LUNG. LOWER  
 CC LEVELS ARE FOUND IN THE LIVER, SPLEEN AND HEART, BUT NOT IN THE  
 CC BRAIN.  
 CC -1- INDUCTION: BY RETINOIC ACID.  
 CC -1- SIMILARITY: BELONGS TO THE TRANSGLUTAMINASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X60686; CAA43097.1;  
 CC HSSP; P00488; 1FIE.  
 CC InterPro; IPR001102;  
 CC InterPro; IPR002931;  
 CC Pfam; PF01841; Transglut\_core; 1.  
 CC Pfam; PF00927; Transglutamin\_C; 1.  
 CC Pfam; PF00868; Transglutamin\_N; 1.  
 CC PROSITE; PS00547; TRANSGLUTAMINASES; 1.  
 KW Transferase; Acyltransferase; Calcium-binding.  
 FT ACT\_SITE 277 277 BY SIMILARITY.  
 SQ SEQUENCE 687 AA; 77112 MW; 7BBA00F15E779944 CRC64;

```

Query Match      51.3%; Score 40; DB 1; Length 687;
Best Local Similarity 46.7%; Pred. No. 35;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 ADEEEQQALSSQMGF 16
DB 152 SDEERQEVLTQOGF 166

RESULT 5
ACTU_DROME STANDARD; PRT; 398 AA.
AC ACTU_DROME Q9VXQ9;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ACTIN-LIKE PROTEIN 135.
GN ACTR135 OR ACRP OR G11678.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=94141930; PubMed=8308899;
RA Frankel S., Heintzelman M.B., Artavanis-Tsakonas S., Mooseker M.S.;
RT "Identification of a divergent actin-related protein in Drosophila.";
RL J. Mol. Biol. 235:1351-1356(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Borkova D., Bouch M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.A., Nixon K., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu C., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- SIMILARITY: BELONGS TO THE ACTIN FAMILY. ARP6 SUBFAMILY.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: L25314; AAA17685.1;
EMBL: AE003500; AAF48499.1;
DR FlyBase; FBgn0011741; Actr135.
DR InterPro; IPR000279;
DR Pfam; PF00022; actin; 2.
KW Structural protein; Cytoskeleton.
SQ SEQUENCE 398 AA; 45414 MW; 71905627B3F0FC05 CRC64;

Query Match      50.0%; Score 39; DB 1; Length 398;
Best Local Similarity 72.7%; Pred. No. 29;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RADEEQQQALS 11
DB 264 REDEEQQQMVS 274

RESULT 6
GVGL_HALN1 STANDARD; PRT; 83 AA.
AC P24371; Q9H122;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GVPG PROTEIN 1.
GN (GVPG11 OR GVPG OR VNG5026G) AND (GVPG12 OR VNG5025G).
OS Halobacterium sp. (strain NRC-1), and
OS Halobacterium sp. (strain NRC-817).
OG Plasmid pNRC100, plasmid pNRC200, and Plasmid pH11.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091, 148370;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRC-1; PLASMID=pNRC100;
RX MEDLINE=91323716; PubMed=1864501;
RA Jones J.G., Young D.C., Dassarma S.;
RT "Structure and organization of the gas vesicle gene cluster on the
RT Halobacterium halobium plasmid pNRC100.";
RL Gene 102:117-122(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRC-1; PLASMID=pNRC100;
RX MEDLINE=99063795; PubMed=9847077;
RA Ng W.V., Clufo S.A., Smith T.M., Bungarner R.E., Baskin D., Faust J.,
RA Hall B., Loretz C., Seto J., Slagel J., Hood L., Dassarma S.;
RT "Snapshot of a large dynamic replicon in a halophilic archaeon:
RT megaplasmid or minichromosome?";
RL Genome Res. 8:1131-1141(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NRC-1; PLASMID=pNRC200;
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitthauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

```

[4]  
SEQUENCE FROM N.A.  
STRAIN-NRC-817; PLASMID=PHH1;  
MEDLINE=92065812; PubMed=1956294;  
Horne M., Englert C., Wimmer C., Pfeifer F.;  
"A DNA region of 9 kbp contains all genes necessary for gas vesicle  
synthesis in halophilic archaeobacteria";  
Mol. Microbiol. 5:1159-1174(1991).  
-!- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS  
VESICLE SYNTHESIS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M58557; AAA98192.1; -  
DR EMBL; AF016485; AAC82805.1; -  
DR EMBL; AE005141; AAG20722.1; -  
DR EMBL; X55648; CAA39174.1; -  
DR PIR; JQ1124; JQ1124.  
DR PIR; S15186; S15186.  
KW Gas vesicle; Plasmid.  
SQ SEQUENCE 83 AA; 10014 MW; 4FCPE7B4EF43C792 CRC64;  
  
Query Match 49.4%; Score 38.5; DB 1; Length 83;  
Best Local Similarity 58.8%; Pred. No. 6.7;  
Matches 10; Conservative 2; Mismatches 2; Indels 3; Gaps 1;  
  
Qy 1 RADEEQ---QALSSQM 14  
Db 49 RSDEEQKQKQALSSQ 65  
  
RESULT 7  
TFA\_BPMU STANDARD; PRT; 175 AA.  
AC Q9RIU9;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PROBABLE TAIL FIBER ASSEMBLY PROTEIN (GFP).  
GN U OR 50.  
OS Bacteriophage Mu.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.  
OX NCBI\_TaxID=10677;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Morgan G., Hatfull G., Hendrix R.;  
"Genome of Bacteriophage Mu and comparison with the Haemophilus  
influenzae Mu-like prophage Flumu.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CHAPERONE INVOLVED IN TAIL FIBER ASSEMBLY (BY  
SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE TFA FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF083977; AAF01128.1; -  
DR EMBL; AF083977; AAF01128.1; -  
DR EMBL; AF083977; AAF01128.1; -  
KW Chaperone.  
SQ SEQUENCE 175 AA; 20310 MW; 71A0EBA3294FCCD1 CRC64;

Query Match 48.7%; Score 38; DB 1; Length 175;  
Best Local Similarity 66.7%; Pred. No. 18;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 2 ADEEQQALSSQ 13  
Db 108 ADEEQQALSSQ 119  
  
RESULT 8  
TFA\_BPP2 STANDARD; PRT; 175 AA.  
AC P26699;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PROBABLE TAIL FIBER ASSEMBLY PROTEIN (GFP).  
GN G.  
OS Bacteriophage P2.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.  
OX P2-like Viruses.  
OX NCBI\_TaxID=10679;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92165720; PubMed=1531648;  
RA Haggaard-Ljungquist E., Halling C., Calendar R.;  
"DNA sequences of the tail fiber genes of bacteriophage P2: evidence  
for horizontal transfer of tail fiber genes among unrelated  
bacteriophages";  
J. Bacteriol. 174:1462-1477(1992).  
CC -!- FUNCTION: ACTS CATALYTICALLY IN THE FORMATION OF TAIL PROTEIN  
DIMERS.  
CC -!- SIMILARITY: BELONGS TO THE TFA FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF063097; AAD03287.1; -  
DR PIR; C42291; C42291.  
KW Chaperone.  
SQ SEQUENCE 175 AA; 20287 MW; A2FFA76948260F17 CRC64;  
  
Query Match 48.7%; Score 38; DB 1; Length 175;  
Best Local Similarity 66.7%; Pred. No. 18;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 2 ADEEQQALSSQ 13  
Db 108 ADEEQQALSSQ 119  
  
RESULT 9  
DN43\_YEAST STANDARD; PRT; 571 AA.  
AC P32354;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE DN43 PROTEIN.  
GN DN43 OR MCM10 OR YIL150C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92383947; PubMed=1514326;





DR MIM; 150310; -  
 DR MIM; 226700; -  
 DR InterPro: IPR000561; -  
 DR InterPro: IPR001886; -  
 DR InterPro: IPR002049; -  
 DR DR PF00053; laminin\_EGF; 6.  
 DR Pfam: PF00055; laminin\_Nterm; 1.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR PROSITE; PS00022; EGF\_1; 5.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 5.  
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal;  
 KW Disease mutation.  
 FT SIGNAL 1 17  
 FT CHAIN 18 1172  
 FT DOMAIN 18 249  
 FT DOMAIN 250 578  
 FT  
 FT DOMAIN 250 315  
 FT DOMAIN 316 378  
 FT DOMAIN 379 430  
 FT DOMAIN 431 480  
 FT DOMAIN 481 533  
 FT DOMAIN 534 578  
 FT DOMAIN 579 785  
 FT DOMAIN 786 816  
 FT DOMAIN 817 1170  
 FT DOMAIN 723 757  
 FT DOMAIN 831 884  
 FT DOMAIN 948 1133  
 FT DISULFID 252 279  
 FT DISULFID 281 290  
 FT DISULFID 293 313  
 FT DISULFID 316 325  
 FT DISULFID 318 343  
 FT DISULFID 346 355  
 FT DISULFID 358 376  
 FT DISULFID 379 392  
 FT DISULFID 381 399  
 FT DISULFID 401 410  
 FT DISULFID 413 428  
 FT DISULFID 431 444  
 FT DISULFID 433 451  
 FT DISULFID 453 462  
 FT DISULFID 465 478  
 FT DISULFID 481 493  
 FT DISULFID 483 500  
 FT DISULFID 502 511  
 FT DISULFID 519 531  
 FT DISULFID 534 546  
 FT DISULFID 536 553  
 FT DISULFID 555 564  
 FT DISULFID 567 578  
 FT DISULFID 581 581  
 FT DISULFID 584 584  
 FT DISULFID 1171 1171  
 FT CARBOHYD 220 220  
 FT CARBOHYD 604 604  
 FT CARBOHYD 810 810  
 FT VARIANT 210 210  
 FT  
 FT VARIANT 679 679  
 FT  
 FT CONFLICT 124 124  
 FT CONFLICT 269 269  
 FT CONFLICT 388 388  
 FT CONFLICT 426 427  
 FT CONFLICT 440 441  
 FT CONFLICT 603 603  
 FT CONFLICT 815 815  
 FT SEQUENCE 1172 AA; 129572 MW; 61BC1A60BBD4FA05 CRC64;

Query Match 48.7%; Score 38; DB 1; Length 1172;  
 Best Local Similarity 53.3%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 ADEEQQALSSQMGF 16  
 DB 1064 AEGASEQALSAQEGF 1078  
 ID: |||||: ||  
 RESULT 11  
 NEST\_RAT  
 ID NEST\_RAT STANDARD; PRT; 1805 AA.  
 AC P21263;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE NESTIN.  
 GN NEST.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90150286; PubMed=1689217;  
 RA Lendahl U., Zimmerman L.B., McKay R.D.G.;  
 RT "CNS stem cells express a new class of intermediate filament  
 protein.";  
 RL Cell 60:585-595(1990).  
 CC -!- TISSUE SPECIFICITY: CNS STEM CELLS.  
 CC -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN  
 CC -!- IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.  
 CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M34384; AAA41685.1; -  
 DR PIR; A34736; A34736.  
 DR InterPro: IPR001664; -  
 DR Pfam: PF00038; filament; 2.  
 DR PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Heptad repeat pattern;  
 FT Neutrone.  
 FT DOMAIN 1 7  
 FT DOMAIN 8 314  
 FT DOMAIN 315 1805  
 FT DOMAIN 8 43  
 FT DOMAIN 44 55  
 FT DOMAIN 56 151  
 FT DOMAIN 152 174  
 FT DOMAIN 175 193  
 FT DOMAIN 194 196  
 FT DOMAIN 197 314  
 FT SEQUENCE 1805 AA; 198744 MW; B40EE14717E0998D CRC64;

Query Match 48.7%; Score 38; DB 1; Length 1805;  
 Best Local Similarity 58.3%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 EEOQQALSSQMG 15  
 DB 30 EEQQLLSAELG 41  
 ID: |||||: ||

```

RESULT 12
CBPA_PIG          STANDARD;          PRT;          94 AA.
ID CBPA_PIG
AC P09954;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CARBOXYPEPTIDASE A (EC 3.4.17.1) (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE.
RX MEDLINE=87100171; PubMed=3801014;
RA Vendrell J., Aviles F.X., Genesca E., San Segundo B., Soriano F.,
RA Mendez E.;
RT "Primary structure of the activation segment of procarboxypeptidase A
RT from porcine pancreas.";
RL Biochem. Biophys. Res. Commun. 141:517-523(1986).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2 ANGSTROMS).
RX MEDLINE=92194312; PubMed=1548696;
RA Guasch A., Coll M., Aviles F.X., Huber R.;
RA "Three-dimensional structure of porcine pancreatic
RT procarboxypeptidase A. A comparison of the A and B zymogens and their
RT determinants for inhibition and activation.";
RL J. Mol. Biol. 224:141-157(1992).
CC -1- CATALYTIC ACTIVITY: PEPTIDYL-L-AMINO ACID + H(2)O - PEPTIDE +
CC L-AMINO ACID.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14; ALSO KNOWN AS THE
CC ZINC CARBOXYPEPTIDASE FAMILY
DR PIR; A25833; A25833.
DR PDB; 1PCA: 31-OCT-93.
DR MEROPS; M14.001;
DR InterPro; IPR000834;
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; PARTIAL.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; PARTIAL.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KW 3D-structure.
FT PROPEP          1 94      ACTIVATION PEPTIDE.
FT TURN           5 6
FT STRAND         8 12
FT HELIX          17 26
FT TURN           27 28
FT TURN           30 32
FT HELIX          35 38
FT STRAND         43 44
FT TURN           47 51
FT STRAND         53 65
FT HELIX          66 67
FT TURN           70 74
FT STRAND         77 90
FT HELIX          91 91
FT TURN           94 94
FT NON_TER
SQ SEQUENCE 94 AA; 10516 MW; 0E1DB779C4F475B0 CRC64;

Query Match 47.4%; Score 37; DB 1; Length 94;
Best Local Similarity 63.6%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 DEEQQQALSSQ 13
Db 82 DEEQQMFASQ 92
|||||:|:|
-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN MAINLY BECAUSE

RESULT 13
TPMM_LOCM1
ID TPMM_LOCM1
AC P31816;
DT 01-JUL-1993 (Rel. 26, Created)

```

```

DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TROPOMYOSIN, MUSCLE.
OS Locusta migratoria (Migratory locust).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC Acridomorpha; Acridoidea; Acrididae; Locusta.
OX NCBI_TaxID=7004;
RN [1]
RP SEQUENCE FROM N.A.
RA Krieger J., Raming K., Knipper M., Grau M., Mertens S., Breer H.;
RT "Cloning, sequencing and expression of locust tropomyosin.";
RL Insect Biochem. 20:173-184(1990).
CC -1- FUNCTION: TROPOMYOSIN, IN ASSOCIATION WITH THE TROPONIN COMPLEX,
CC PLAYS A CENTRAL ROLE IN THE CALCIUM DEPENDENT REGULATION OF
CC MUSCLE CONTRACTION.
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
DR PIR; A60364; A60364.
DR HSSP; P01100; 1FOS.
DR InterPro; IPR000533;
DR Pfam; PF00261; Tropomyosin; 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR PROSITE; PS00326; TROPOMYOSIN; 1.
KW Muscle protein; Coiled coil; Repeat.
SQ SEQUENCE 283 AA; 32439 MW; 8C9E25FEFA664C0A CRC64;

Query Match 47.4%; Score 37; DB 1; Length 283;
Best Local Similarity 53.8%; Pred. No. 43;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 ADEEQQQALSSQM 14
Db 136 ADEERMDALENQL 148
|||||:|:|
-1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN MAINLY BECAUSE

RESULT 14
TPML_DROME
ID TPML_DROME          STANDARD;          PRT;          285 AA.
AC P06754;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TROPOMYOSIN I, MUSCLE ISOFORM (TROPOMYOSIN II).
GN TM1 OR TM1I.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87106815; PubMed=3803921;
RA Hanke P.D., Storti R.V.;
RT "Nucleotide sequence of a cDNA clone encoding a Drosophila muscle
RT tropomyosin II isoform.";
RL Gene 45:211-214(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064486; PubMed=3097506;
RA Karlik C.C., Fyrberg E.A.;
RT "Two Drosophila melanogaster tropomyosin genes: structural and
RT functional aspects.";
RL Mol. Cell. Biol. 6:1965-1973(1986).
CC -1- ALTERNATIVE PRODUCTS: DROSOPHILA TROPOMYOSIN 1 GENE CAN PRODUCE
CC FOUR DIFFERENT ISOFORMS BY ALTERNATIVE SPLICING: A MUSCLE FORM,
CC A NON-MUSCLE FORM, AND TWO FUSION PROTEINS (33 AND 34).
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN MAINLY BECAUSE

```



SQ SEQUENCE 386 AA; 40757 MW; 2F5C89D9FACB4550 CRC64;

Query Match 47.4%; Score 37; DB 1; Length 386;  
 Best Local Similarity 61.5%; Pred. No. 60;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 ADEEQQQALSSQM 14  
 |||:| ||: ||  
 Db 304 ADEDQLALALQM 316

Search completed: July 5, 2001, 11:51:46  
 Job time: 505 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:50:52 ; Search time 123.78 Seconds  
(without alignments)  
17.102 Million cell updates/sec

Title: US-09-462-480-13

Perfect score: 78

Sequence: 1 RADEEQQALSSQMGF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_16.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_prodent.\*

12: sp\_unclassified.\*

13: sp\_vertebrate.\*

14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	78	100.0	100	2	Q69739	Q69739 mycobacteri
2	49	62.8	1263	5	Q09971	Q09971 caenorhabdi
3	43	55.1	412	14	O10415	O10415 helicoverpa
4	42	53.8	100	2	O33084	O33084 mycobacteri
5	42	53.8	224	2	O86809	O86809 streptomyce
6	42	53.8	248	4	Q9NVV1	Q9NVV1 homo sapien
7	42	53.8	902	2	Q9I742	Q9I742 pseudomonas
8	41	52.6	270	10	Q9M4L9	Q9M4L9 triticum ae
9	41	52.6	856	5	Q23419	Q23419 caenorhabdi
10	40.5	51.9	258	2	Q9HX16	Q9HX16 pseudomonas
11	40	51.3	239	2	Q9RJV3	Q9RJV3 streptomyce
12	40	51.3	437	2	O33491	O33491 pseudomonas
13	40	51.3	472	5	O96967	O96967 drosophila
14	40	51.3	486	5	Q9VKE2	Q9VKE2 drosophila
15	40	51.3	548	2	O54414	O54414 acinetobact
16	40	51.3	911	2	Q56658	Q56658 vibrio chol
17	39	50.0	355	10	Q9LN35	Q9LN35 arabidopsis
18	39	50.0	682	10	Q39948	Q39948 helianthus
19	39	50.0	827	5	Q19107	Q19107 caenorhabdi

20	39	50.0	838	10	O65525	O65525 arabidopsis
21	39	50.0	864	11	Q60943	Q60943 mus musculu
22	39	50.0	902	5	O46035	O46035 drosophila
23	39	50.0	993	5	Q9VY31	Q9VY31 drosophila
24	39	50.0	1105	4	O60518	O60518 homo sapien
25	39	50.0	1286	5	Q21025	Q21025 caenorhabdi
26	38	48.7	144	5	Q9NAC4	Q9NAC4 caenorhabdi
27	38	48.7	175	2	O57237	O57237 shigella bo
28	38	48.7	198	9	O03936	O03936 bacterioph
29	38	48.7	213	5	Q9V8B3	Q9V8B3 drosophila
30	38	48.7	353	2	Q9FCD8	Q9FCD8 streptomyce
31	38	48.7	369	5	Q9V5V7	Q9V5V7 drosophila
32	38	48.7	381	5	Q09585	Q09585 caenorhabdi
33	38	48.7	386	4	Q9UTL1	Q9UTL1 homo sapien
34	38	48.7	397	2	Q9P9P7	Q9P9P7 xylella fas
35	38	48.7	412	14	O10610	O10610 helicoverpa
36	38	48.7	529	2	P74630	P74630 synechocyst
37	38	48.7	564	2	Q9I2B7	Q9I2B7 pseudomonas
38	38	48.7	593	5	Q9VZW9	Q9VZW9 drosophila
39	38	48.7	601	2	Q9I612	Q9I612 pseudomonas
40	38	48.7	768	10	O40988	O40988 phalaenopsi
41	38	48.7	866	4	O43844	O43844 homo sapien
42	38	48.7	872	5	Q90300	Q90300 caenorhabdi
43	38	48.7	1172	4	O14947	O14947 homo sapien
44	38	48.7	1463	5	Q9GY20	Q9GY20 strongyloce
45	38	48.7	1821	11	Q9R0C4	Q9R0C4 mus musculu

#### ALIGNMENTS

RESULT 1  
ID O69739 PRELIMINARY; PRT; 100 AA.  
AC O69739;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE HYPOTHETICAL 10.8 KDA PROTEIN.  
GN LHP OR RV3874 OR MTV027.09.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RT Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;  
RT "Promoter analysis of the M. tuberculosis orf1c gene encoding the  
early secreted antigenic target 6 kDa (ESAT-6).";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF022120; CAAL1966.1; -;  
DR EMBL: AF004671; AAC83445.1; -;  
DR Tuberculist; RV3874; -;  
KW Hypothetical protein.  
SQ SEQUENCE 100 AA; 10794 MW; 285F4FC96F55D194 CRC64;

Query Match 100.0%; Score 78; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.1e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RADEEQOQALSSQMGF 16  
| | | | | | | | | | | | | | | |  
DB 85 RADEEQOQALSSQMGF 100

RESULT 2  
Q09971 ID Q09971 PRELIMINARY; PRT; 1263 AA.  
AC Q09971;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE HYPOTHEICAL 145.0 KDA PROTEIN C14F5.3 IN CHROMOSOME X.  
GN C14F5.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Minx P.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U29082; AAA68402.1; -;  
DR WormPep: C14F5.3; CE01782.  
DR InterPro: IPR001978; -;  
DR Pfam: PF00992; Troponin; 1.  
KW Hypothetical protein.  
FT DOMAIN 35 38 POLY-SER.  
FT DOMAIN 176 181 POLY-SER.  
FT DOMAIN 253 258 POLY-SER.  
FT DOMAIN 920 987 ARG/GLU-RICH.  
FT DOMAIN 926 933 POLY-GLU.  
FT DOMAIN 969 972 POLY-GLU.  
FT DOMAIN 1163 1166 POLY-PRO.  
SQ SEQUENCE 1263 AA; 145003 MW; 1C17172C2B3D6EC0 CRC64;

Query Match 62.8%; Score 49; DB 5; Length 1263;  
Best Local Similarity 60.0%; Pred. No. 2.6;  
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 ADEEQOQALSSQMGF 16  
| | | | | | | | | | | | | | | |  
DB 28 SDEEQYSSSSQLGF 42

RESULT 3  
O10415 ID O10415 PRELIMINARY; PRT; 412 AA.  
AC O10415;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE ORF1236.  
OS Helicoverpa armigera nuclear polyhedrosis virus.  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=51313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhang C., Wang G., Hu C., Wu X.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U95055; AAB53737.1; -;  
DR InterPro: IPR003124; -;  
DR Pfam: PF02205; WH2; 1.  
DR SMART: SM00246; WH2; 1.  
SQ SEQUENCE 412 AA; 45652 MW; CA506F8893863386 CRC64;

Query Match 55.1%; Score 43; DB 14; Length 412;  
Best Local Similarity 57.1%; Pred. No. 9.9;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RADEEQOQALSSOM 14  
| | | | | | | | | | | | | | | |  
DB 256 RTDDEQOQRASSEL 269

RESULT 4  
O33084 ID O33084 PRELIMINARY; PRT; 100 AA.  
AC O33084;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE SIMILARITY TO SMALL.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Eiglmeier K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE-93188700; PubMed-8446027;  
RA Eiglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;  
RT "Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae";  
RL Mol. Microbiol. 7:197-206(1993).  
DR EMBL: Y14967; CAA75210.1; -;  
SQ SEQUENCE 100 AA; 10964 MW; 460EE12F876EC383 CRC64;

Query Match 53.8%; Score 42; DB 2; Length 100;  
Best Local Similarity 50.0%; Pred. No. 3.5;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 RADEEQOQALSSQMGF 16  
| | | | | | | | | | | | | | | |  
DB 85 KTDDEANQLSSKNMF 100

RESULT 5  
O86809 ID O86809 PRELIMINARY; PRT; 224 AA.  
AC O86809;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE TWO-COMPONENT REGULATOR.  
GN SC7C7.04.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Harris D., Taylor K.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RX MEDLINE-97000351; PubMed-8843436;

RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RT Mol. Microbiol. 21:77-96(1996).  
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER  
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.  
DR EMBL; AL031031; CAAL19850.1; -  
DR HSSP; P41789; INTR.  
DR InterPro: IPR001789; -  
DR Pfam; PF000072; response\_reg; 1.  
DR SMART; SM00448; REC; 1.  
KW Phosphorylation; Sensory transduction.  
SQ SEQUENCE 224 AA; 24101 MW; 2CD9EF5DC4A39B2B CRC64;

Query Match 53.8%; Score 42; DB 2; Length 224;  
Best Local Similarity 64.3%; Pred. No. 8;

Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 ADEEQQALSSQMG 15

| : : : | : : |

Db 170 AVEEQAEALSKQLG 183

RESULT 6

Q9NVV1 PRELIMINARY; PRT; 248 AA.

AC Q9NVV1;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE CDNA FLJ10491 FIS. CLONE NT2RP2000239.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tsogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosolri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masubo Y.,  
RA Ninomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK001353; BAA91644.1; -  
SQ SEQUENCE 248 AA; 27462 MW; E9802B0415EDF62A CRC64;

Query Match 53.8%; Score 42; DB 4; Length 248;  
Best Local Similarity 61.5%; Pred. No. 8.9;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 EEOQQALSSQMGF 16

| : : : | : : |

Db 227 EERKQALSDRQGF 239

RESULT 7

Q91742 PRELIMINARY; PRT; 902 AA.

AC Q91742;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE PROBABLE CLPA/B-TYPE CHAPERONE.  
GN PA0090.

OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004447; AAG03480.1; -  
DR InterPro: IPR001270; -  
DR InterPro: IPR001939; -  
DR Pfam; PF00004; AAA; 1.  
DR PRINTS; PR00300; CLPPROTEASEA.  
DR PROSITE; PS00870; CLPAB\_1; 1.  
SQ SEQUENCE 902 AA; 98737 MW; DB45A33B091C0071 CRC64;

Query Match 53.8%; Score 42; DB 2; Length 902;

Best Local Similarity 57.1%; Pred. No. 33;

Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RADEEQQALSSQMG 14

| : : : | : : |

Db 532 RLDEEQRQALHARL 545

RESULT 8

Q9M4L9 PRELIMINARY; PRT; 270 AA.

AC Q9M4L9;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE ALPHA-GLIADIN.  
OS Triticum aestivum (Wheat).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Triticeae;  
OC Triticum.  
OX NCBI\_TaxID=4565;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=MJOELNER; TISSUE-ENDOSPERM;  
RA Arentz-Hansen E.H., McAdam S.N., Molberg O., Kristiansen C.,  
RA Sollid L.M.;  
RT "Production of a panel of recombinant gliadins for the  
RT characterisation of T cell reactivity in coeliac disease.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ133609; CAB76961.1; -  
DR InterPro: IPR001954; -  
DR InterPro: IPR003612; -  
DR PRINTS; PR00208; GLIADGLUTEN.  
DR SMART; SM00499; AAL; 1.  
SQ SEQUENCE 270 AA; 31491 MW; 1DB4B6528EADF55 CRC64;

Query Match 52.6%; Score 41; DB 10; Length 270;

Best Local Similarity 61.5%; Pred. No. 15;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 EEOQQALSSQMGF 16

| : : : | : : |

Db 187 OKOQQALSSQVSF 199

RESULT 9

Q23419 PRELIMINARY; PRT; 856 AA.







Query Match 51.3%; Score 40; DB 5; Length 486;  
 Best Local Similarity 66.7%; Pred. No. 41;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 RADEEQQALSS 12  
 | |||:|:|:  
 Db 146 RLDEQQQORLSA 157

RESULT 15  
 O54414 PRELIMINARY; PRT; 548 AA.  
 AC O54414;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE MALONATE DECARBOXYLASE ALPHA SUBUNIT.  
 GN MDCA.  
 OS Acinetobacter calcoaceticus.  
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
 OC Acinetobacter.  
 OX NCBI\_TaxID=471;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98041639; PubMed=9375791;  
 RA KOO J.H., Jung S.B., Byun H.S., Kim Y.S.;  
 RT "Cloning and sequencing of genes encoding malonate decarboxylase in  
 RT Acinetobacter calcoaceticus.";  
 RL Biochim. Biophys. Acta 1354:49-54 (1997).  
 DR EMBL; AF121266; AAB97627.1; -.  
 DR InterPro; IPR002086; -.  
 DR InterPro; IPR002106; -.  
 DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
 DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
 SQ SEQUENCE 548 AA; 60899 MW; C712AEC6CD8A52A4 CRC64;

Query Match 51.3%; Score 40; DB 2; Length 548;  
 Best Local Similarity 43.8%; Pred. No. 46;  
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 RADEEQQALSSQMGF 16  
 | |||:|:|:  
 Db 469 RTDEEQAIRGVAGY 484

Search completed: July 5, 2001, 11:50:53  
 Job time: 476 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:45:44 ; Search time 130.35 Seconds  
(without alignments)  
13.022 Million cell updates/sec

Title: US-09-462-480-12  
Perfect score: 137  
Sequence: 1 STNIRAGVGYRADEQQALSSQMGF 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues  
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0501.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	137	100.0	28	20	AAW0712	M. tuberculosis LH
2	137	100.0	42	20	AAW0707	M. tuberculosis LH
3	137	100.0	80	18	AAW32454	Mycobacterium tube
4	137	100.0	80	18	AAW32386	Mycobacterium tube
5	137	100.0	80	19	AAW81707	M. tuberculosis im
6	137	100.0	80	19	AAW64340	Mycobacterium tube
7	137	100.0	80	20	AAW39137	M. tuberculosis an
8	137	100.0	80	20	AAW38994	M. tuberculosis re
9	137	100.0	95	18	AAW32444	Mycobacterium tube
10	137	100.0	95	18	AAW32376	Mycobacterium tube
11	137	100.0	95	19	AAW81747	M. tuberculosis im

12	137	100.0	95	19	AAW64321	Mycobacterium tube
13	137	100.0	95	20	AAW32097	Mycobacterium tube
14	137	100.0	95	20	AAW39118	M. tuberculosis an
15	137	100.0	95	20	AAW38981	M. tuberculosis re
16	137	100.0	100	19	AAW81706	Mycobacterium im
17	137	100.0	100	19	AAW64339	Mycobacterium tube
18	137	100.0	100	20	AAW39136	M. tuberculosis an
19	137	100.0	100	20	AAW38993	M. tuberculosis re
20	137	100.0	100	20	AAW03705	M. tuberculosis LH
21	137	100.0	100	22	AAW35218	M. tuberculosis RV3
22	137	100.0	100	22	AAW19845	Mycobacterium tube
23	137	100.0	802	19	AAW81746	M. tuberculosis fu
24	137	100.0	802	19	AAW64379	Mycobacterium anti
25	137	100.0	802	20	AAW32063	Mycobacterium tube
26	137	100.0	802	20	AAW39224	M. tuberculosis fu
27	137	100.0	802	20	AAW39176	M. tuberculosis fu
28	137	100.0	802	20	AAW39081	M. tuberculosis fu
29	137	100.0	802	20	AAW39033	M. tuberculosis fu
30	122	89.1	28	18	AAW32460	Mycobacterium tube
31	122	89.1	28	19	AAW81698	M. tuberculosis im
32	122	89.1	28	20	AAW39128	M. tuberculosis an
33	78	56.9	16	20	AAW03713	M. tuberculosis LH
34	64	46.7	27	18	AAW32459	Mycobacterium tube
35	64	46.7	27	19	AAW81697	M. tuberculosis im
36	64	46.7	27	20	AAW39127	M. tuberculosis an
37	53	38.7	276	20	AAW00153	Enterococcus faeca
38	53	38.7	303	20	AAW00152	Enterococcus faeca
39	46.5	33.9	1325	18	AAW19540	Male-enhanced anti
40	46.5	33.9	1325	20	AAW94391	Mouse male enhance
41	46	33.6	431	20	AAW35378	Chlamydia pneumonia
42	46	33.6	493	21	AAW44495	Bacillus agaradher
43	46	33.6	493	21	AAW54124	Amino acid sequenc
44	45.5	33.2	200	21	AAW69370	A 5'-OT EST (oxyto
45	45	32.8	691	11	AAW04711	Sequence of guinea

## ALIGNMENTS

RESULT	1
AAW03712	AAW03712 standard; Protein; 28 AA.
ID	AAW03712 standard; Protein; 28 AA.
XX	AAW03712;
AC	AAW03712;
DT	07-JUN-1999 (first entry)
XX	XX
DE	M. tuberculosis LHP polypeptide antigenic fragment.
XX	XX
KW	ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;
KW	immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;
KW	immune response.
XX	XX
OS	Mycobacterium tuberculosis.
XX	XX
PN	WO9904005-A1.
XX	XX
PD	28-JAN-1999.
XX	XX
PF	16-JUL-1998; 98WO-IB01091.
XX	XX
PR	16-JUL-1997; 97US-0052631.
XX	XX
PA	(INSP ) INST PASTEUR.
PA	(STAT-) STATENS SERUM INST.
XX	XX
PI	Andersen P, Berthet F, Gicquel B, Rasmussen PB;
XX	XX
DR	WPI; 1999-132249/11.
XX	XX
PT	New nucleic acid containing regulator and LHP gene of Mycobacterium
PT	tuberculosis - useful in vaccines, for diagnosis, and for expression
PT	of heterologous proteins



CC non-specific adjuvant, to induce an immune response against  
CC M.tuberculosis (for treatment or prevention).

XX  
SQ Sequence 80 AA;

Query Match 100.0%; Score 137; DB 18; Length 80;  
Best Local Similarity 100.0%; Pred. No. 2e-14;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STNIRAGVQYSRADEEQQALSSQMGF 28  
Db 53 stniragvqysradeeqqalssqmgf 80

RESULT 4  
AAW32386  
ID AAW32386 standard; Protein; 80 AA.

XX  
AC AAW32386;

XX  
DT 13-JAN-1998 (first entry)

XX  
DE Mycobacterium tuberculosis antigen Tb38-1F3.

XX  
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

XX  
KW skin testing; M.tuberculosis.

XX  
OS Mycobacterium tuberculosis.

XX  
PN WO9709429-A2.

XX  
PD 13-MAR-1997.

XX  
PF 30-AUG-1996; 96WO-US14675.

XX  
PR 12-JUL-1996; 96US-0680573.

XX  
PR 01-SEP-1995; 95US-0523435.

XX  
PR 22-SEP-1995; 95US-0532136.

XX  
PR 22-MAR-1996; 96US-0620280.

XX  
PR 05-JUN-1996; 96US-0658800.

XX  
PA (CORI-) CORIXA CORP.

XX  
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

XX  
PI Twardzik DR, Vedvick TH;

XX  
DR WPI; 1997-192904/17.

XX  
DR N-PSDB; AAT91460.

XX  
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
PT - useful for diagnosis of M. tuberculosis infection

XX  
PS Example 3; Page 163; 190pp; English.

XX  
CC A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb38-1F3. The immunogenic polypeptide can be used to diagnose  
CC M.tuberculosis infection by forming complexes with specific  
CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
CC polypeptide can be used as diagnostic primers or probes and agents  
CC that bind to the antigen, especially monoclonal antibodies or  
CC equivalent polyclonal antibodies, are also used for diagnosis.

XX  
SQ Sequence 80 AA;

Query Match 100.0%; Score 137; DB 18; Length 80;  
Best Local Similarity 100.0%; Pred. No. 2e-14;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STNIRAGVQYSRADEEQQALSSQMGF 28  
Db 53 stniragvqysradeeqqalssqmgf 80

RESULT 5

AAW81707

ID AAW81707 standard; Protein; 80 AA.

XX  
AC AAW81707;

XX  
DT 27-JAN-1999 (first entry)

XX  
DE M. tuberculosis immunogenic polypeptide Tb38-1F3.

XX  
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis.

XX  
OS Mycobacterium tuberculosis.

XX  
PN WO9816646-A2.

XX  
PD 23-APR-1998.

XX  
PF 07-OCT-1997; 97WO-US18293.

XX  
PR 13-MAR-1997; 97US-0818112.

XX  
PR 11-OCT-1996; 96US-0730510.

XX  
PA (CORI-) CORIXA CORP.

XX  
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX  
DR WPI; 1998-261042/23.

XX  
DR N-PSDB; AAV64509.

XX  
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis

XX  
PS Example 3B; Page 139-140; 230pp; English.

XX  
CC This sequence represents an immunogenic portion of a soluble  
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
CC for inducing protective immunity against tuberculosis (TB). This  
CC sequence can be formulated into vaccines and/or pharmaceutical  
CC compositions for immunising against M. tuberculosis infection or may  
CC be used for the diagnosis of tuberculosis.

XX  
SQ Sequence 80 AA;

Query Match 100.0%; Score 137; DB 19; Length 80;  
Best Local Similarity 100.0%; Pred. No. 2e-14;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STNIRAGVQYSRADEEQQALSSQMGF 28  
Db 53 stniragvqysradeeqqalssqmgf 80

RESULT 6

AAW64340

ID AAW64340 standard; Protein; 80 AA.

XX  
AC AAW64340;

XX  
DT 09-NOV-1998 (first entry)

XX  
DE Mycobacterium tuberculosis antigen Tb38-1F3.

XX  
KW Tuberculosis; infection; diagnosis; antigen; Tb38-1F3.

```

XX OS Mycobacterium tuberculosis strain H37Rv.
XX PN WO9816645-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18214.
XX PR 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1998-251292/22.
XX DR N-PSDB; AAV44400.
XX CC New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX PT and diagnosis of tuberculosis
XX FS Example 3; Page 146; 250pp; English.
XX CC This polypeptide comprises Mycobacterium tuberculosis antigen
XX CC Tb38-1F3. A DNA molecule (see AAV44400) coding for Tb38-1F3 was
XX CC isolated from a M. tuberculosis strain H37Rv genomic library. The
XX CC invention relates to compositions and methods for diagnosing
XX CC tuberculosis. It provides polypeptides (see AAW64291-W64379)
XX CC comprising an antigenic portion of a soluble M. tuberculosis
XX CC antigen, or an immunogenic portion of an M. tuberculosis
XX CC as well as DNA sequences encoding such polypeptides, recombinant
XX CC expression vectors and transformed or transfected host cells. Also
XX CC claimed are methods and diagnostic kits for detecting M.
XX CC tuberculosis infection in a patient using these polypeptides,
XX CC antibodies or oligonucleotide probes and primers, for the diagnosis
XX CC of tuberculosis.
XX SQ Sequence 80 AA:

Query Match 100.0%; Score 137; DB 19; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STNTRQAGVQYSRADEEQQALSSQMGP 28
Db 53 stntrqagvqysradeeqqalssqgmfp 80

RESULT 7
AAAY39137
ID AAY39137 standard; Protein; 80 AA.
XX AC AAY39137;
XX DT 05-NOV-1999 (first entry)
XX DE M. tuberculosis antigen Tb38-1F3 amino acid sequence.
XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;
XX KW immune response; skin test.
XX OS Mycobacterium tuberculosis.
XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US03268.

```

```

XX OS 05-MAY-1998; 98US-0072967.
XX PR 18-FEB-1998; 98US-0025197.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1999-527409/44.
XX DR N-PSDB; AAZ19310.
XX CC New antigens from Mycobacterium tuberculosis useful in diagnostic
XX PT skin tests and protective or therapeutic vaccines or compositions
XX PS Example 3; Page 134-135; 299pp; English.
XX CC The present invention describes polypeptides comprising an immunogenic
XX CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
XX CC are vaccines and fusion protein containing M. tuberculosis Ag's.
XX CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
XX CC other polypeptides fragments, can be used in pharmaceutical compositions
XX CC or vaccines to generate a protective or therapeutic immune response to
XX CC M. tuberculosis and as reagents in skin tests for diagnosis of
XX CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
XX CC by, T, B or natural killer cells and/or macrophages in
XX CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
XX CC AAY39225 are used in the exemplification of the present invention.
XX SQ Sequence 80 AA:

Query Match 100.0%; Score 137; DB 20; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e-14;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STNTRQAGVQYSRADEEQQALSSQMGP 28
Db 53 stntrqagvqysradeeqqalssqgmfp 80

RESULT 8
AAAY38994
ID AAY38994 standard; Protein; 80 AA.
XX AC AAY38994;
XX DT 05-NOV-1999 (first entry)
XX DE M. tuberculosis recombinant antigen protein Tb38-1F3.
XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;
XX KW vaccine; immunity.
XX OS Mycobacterium tuberculosis.
XX PN WO9942118-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US03265.
XX KW 05-MAY-1998; 98US-0072596.
XX PR 18-FEB-1998; 98US-0024753.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1999-527416/44.
XX DR N-PSDB; AAZ19098.
XX PF 17-FEB-1999; 99WO-US03268.

```

PT New polypeptide comprising antigenic portions of M. tuberculosis  
XX  
PS Example 3; Page 180; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding  
CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
CC polypeptides are useful for detecting M. tuberculosis infection in a  
CC biological sample by detecting antibodies which bind with the  
CC polypeptides, and are useful as vaccines for immunizing against  
CC M. tuberculosis infection. The new detection methods are needed as  
CC current vaccination strategies do not provide 100% immunity.  
XX  
SQ Sequence 80 AA;

Query Match 100.0%; Score 137; DB 20; Length 80;  
Best Local Similarity 100.0%; Pred. No. 2e-14;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 STNIRQAGVOYSRADDEEQQALSSQMGF 28  
|||||  
DB 53 stnirgagvgsradeeqqalssqmgf 80

RESULT 9  
AAW32444  
ID AAW32444 standard; Protein; 95 AA.  
XX  
AC AAW32444;  
XX  
XX 09-JAN-1998 (first entry)  
DT  
XX  
DE Mycobacterium tuberculosis antigen Tb38-1.  
XX  
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO9709428-A2.  
XX  
PD 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.  
XX  
XX 12-JUL-1996; 96US-0680574.  
XX  
XX 01-SEP-1995; 95US-0523436.  
XX  
XX 22-SEP-1995; 95US-0533634.  
XX  
XX 05-JUN-1996; 96US-0620874.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
PI Twardzik DR, Vedvick TH;  
XX  
XX WPI; 1997-192903/17.  
XX  
XX N-PSDB; AAT91509.  
XX  
XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
PT useful in vaccines for prevention or treatment of tuberculosis, also  
PT for diagnosis  
XX  
XX Example 3; Page 124; 168pp; English;

XX A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb38-1. The immunogenic protein, and fusion proteins  
CC containing one or more of the proteins or one of the proteins plus  
CC ESAT-6, are useful in vaccines, preferably when formulated with a  
CC non-specific adjuvant, to induce an immune response against

CC M.tuberculosis (for treatment or prevention).  
XX  
SQ Sequence 95 AA;

Query Match 100.0%; Score 137; DB 18; Length 95;  
Best Local Similarity 100.0%; Pred. No. 2.4e-14;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 STNIRQAGVOYSRADDEEQQALSSQMGF 28  
|||||  
DB 68 stnirgagvgsradeeqqalssqmgf 95

RESULT 10  
AAW32376  
ID AAW32376 standard; Protein; 95 AA.

XX  
AC AAW32376;  
XX  
XX 13-JAN-1998 (first entry)  
DT  
XX  
DE Mycobacterium tuberculosis antigen Tb38-1.  
XX  
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO9709429-A2.  
XX  
PD 13-MAR-1997.  
XX  
XX 30-AUG-1996; 96WO-US14675.  
XX  
XX 12-JUL-1996; 96US-0680573.  
XX  
XX 01-SEP-1995; 95US-0523435.  
XX  
XX 22-SEP-1995; 95US-0532136.  
XX  
XX 22-MAR-1996; 96US-0620280.  
XX  
XX 05-JUN-1996; 96US-0658800.

XX (CORI-) CORIXA CORP.  
XX  
XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
PI Twardzik DR, Vedvick TH;  
XX  
XX WPI; 1997-192904/17.  
XX  
XX N-PSDB; AAT91445.  
XX  
XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
PT - useful for diagnosis of M. tuberculosis infection  
XX  
XX Example 3; Page 136; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose  
CC M.tuberculosis infection by forming complexes with specific  
CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
CC polypeptide can be used as diagnostic primers or probes and agents  
CC that bind to the antigen, especially monoclonal antibodies or  
CC equivalent polyclonal antibodies, are also used for diagnosis.

XX Sequence 95 AA;

Query Match 100.0%; Score 137; DB 18; Length 95;  
Best Local Similarity 100.0%; Pred. No. 2.4e-14;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 STNIRQAGVOYSRADDEEQQALSSQMGF 28

Db 68 stnirgagvysradeeqqalssqmgf 95

## RESULT 11

AAW81747  
ID AAW81747 standard; Protein; 95 AA.

XX  
AC AAW81747;

XX  
DT 27-JAN-1999 (first entry)

XX  
DE M. tuberculosis immunogenic polypeptide Tb38-1.

XX  
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis.

XX  
OS Mycobacterium tuberculosis.

XX  
PN WO9816646-A2.

XX  
PD 23-APR-1998.

XX  
PF 07-OCT-1997; 97WO-US18293.

XX  
PR 13-MAR-1997; 97US-0818112.

XX  
PR 11-OCT-1996; 96US-0730510.

XX  
PA (CORI-) CORIXA CORP.

XX  
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX  
DR WPI; 1998-261042/23.

XX  
DR N-PSDB; AAV64491.

XX  
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis

XX  
PS Example 3b; Page 117; 230pp; English.

XX  
CC This sequence represents an immunogenic portion of a soluble

CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
CC for inducing protective immunity against tuberculosis (TB). This sequence  
CC can be formulated into vaccines and/or pharmaceutical compositions for  
CC immunising against M. tuberculosis infection or may be used for the  
CC diagnosis of tuberculosis.

XX  
SQ Sequence 95 AA;

## Query Match

Best Local Similarity 100.0%; Score 137; DB 19; Length 95;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STNTRQAGVQYSRADEEQQALSSQMGF 28

Db 68 stnirgagvysradeeqqalssqmgf 95

## RESULT 12

AAW64321

ID AAW64321 standard; Peptide; 95 AA.

XX  
AC AAW64321;

XX  
DT 09-NOV-1998 (first entry)

XX  
DE Mycobacterium tuberculosis antigen Tb38-1 peptide.

XX  
KW Tuberculosis; infection; diagnosis; antigen; Tb38-1.

XX

OS Mycobacterium tuberculosis strain H37Rv.

XX  
PN WO9816645-A2.

XX  
PD 23-APR-1998.

XX  
PF 07-OCT-1997; 97WO-US18214.

XX  
PR 13-MAR-1997; 97US-0818111.

XX  
PR 11-OCT-1996; 96US-0729622.

XX  
PA (CORI-) CORIXA CORP.

XX  
PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX  
DR WPI; 1998-251292/22.

XX  
DR N-PSDB; AAV44384.

XX  
PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and diagnosis of tuberculosis

XX  
PS Example 3; Page 123; 250pp; English.

XX  
CC This is an antigenic portion of Mycobacterium tuberculosis antigen  
CC Tb38-1. A DNA sequence (see AAV44384) coding for antigen Tb38-1 was  
CC isolated from a M. tuberculosis strain H37Rv expression library  
CC using sera from patients having pulmonary or pleural tuberculosis.  
CC The invention relates to compositions and methods for diagnosing  
CC tuberculosis. It provides polypeptides (see AAW64291-964379)  
CC comprising an antigenic portion of a soluble M. tuberculosis  
CC antigen, or an immunogenic portion of an M. tuberculosis antigen,  
CC as well as DNA sequences encoding such polypeptides, recombinant  
CC expression vectors and transformed or transfected host cells. Also  
CC claimed are methods and diagnostic kits for detecting M.  
CC tuberculosis infection in a patient using these polypeptides,  
CC antibodies or oligonucleotide probes and primers, for the diagnosis  
CC of tuberculosis.

XX  
SQ Sequence 95 AA;

## Query Match

Best Local Similarity 100.0%; Score 137; DB 19; Length 95;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STNTRQAGVQYSRADEEQQALSSQMGF 28

Db 68 stnirgagvysradeeqqalssqmgf 95

## RESULT 13

AAV32097

ID AAY32097 standard; Protein; 95 AA.

XX  
AC AAY32097;

XX  
DT 17-JAN-2000 (first entry)

XX  
DE Mycobacterium tuberculosis antigen Tb38-1.

XX  
KW Tuberculosis; antigen; fusion protein; Tb38-1; diagnosis; therapy;  
KW vaccine; immunogen.

XX  
OS Mycobacterium tuberculosis.

XX  
PN WO9951748-A2.

XX  
PD 14-OCT-1999.

XX  
PF 07-APR-1999; 99WO-US07717.

XX



PR 07-APR-1998; 98US-0056556.  
 PR 30-DEC-1998; 98US-0223040.  
 XX (CORI-) CORIXA CORP.  
 PA Skeiky YAW, Alderson M, Campos-Neto A;  
 PI WPI; 1999-601610/51.  
 XX  
 DR New fusion proteins useful for diagnosis, prevention and treatment of  
 XX tuberculosis -  
 PT Claim 1; Fig 4D; 83pp; English.  
 PS  
 XX This sequence represents the Mycobacterium tuberculosis antigen  
 XX Tb38-1. The invention provides fusion proteins (see AAY32059-71)  
 CC containing at least 2 M. tuberculosis antigens such as Tb38-1, e.g.  
 CC TbF-2 (see AAY32063) and a TbH9-Tb38-1 fusion. The new fusion  
 CC proteins are useful as vaccines for preventing tuberculosis  
 CC (claimed), for diagnosis (via in vitro assays or intradermal skin  
 CC tests for detection of anti-M. tuberculosis antibodies), monitoring  
 CC of disease progression, and treatment of tuberculosis. They are  
 CC more effective immunogens than mixtures of the individual protein  
 CC components.  
 XX  
 XX Sequence 95 AA;  
 SQ

Query Match 100.0%; Score 137; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-14;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STNTRAGVOYSRADDEQQQALSSQMCF 28  
 |||||  
 Db 68 stntrgagvdyrsadeeqqalssqmgf 95

RESULT 14  
 AAY39118  
 ID AAY39118 standard; Protein; 95 AA.  
 XX  
 AC AAY39118;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis antigen Tb38-1 amino acid sequence.  
 XX  
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9942076-A2.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99WO-US03268.  
 XX  
 PR 05-MAY-1998; 98US-0072967.  
 PR 18-FEB-1998; 98US-0025197.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1999-527409/44.  
 XX  
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX  
 PS Example 3; Page 113; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to  
 CC AAY39225 are used in the exemplification of the present invention.  
 XX  
 XX Sequence 95 AA;  
 SQ

Query Match 100.0%; Score 137; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-14;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STNTRAGVOYSRADDEQQQALSSQMCF 28  
 |||||  
 Db 68 stntrgagvdyrsadeeqqalssqmgf 95

RESULT 15  
 AAY38981  
 ID AAY38981 standard; Protein; 95 AA.  
 XX  
 AC AAY38981;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis recombinant antigen protein Tb38-1.  
 XX  
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9942118-A2.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99WO-US03265.  
 XX  
 PR 05-MAY-1998; 98US-0072596.  
 PR 18-FEB-1998; 98US-0024753.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1999-527416/44.  
 DR N-PSDB; AA219082.  
 XX  
 PT New polypeptide comprising antigenic portions of M. tuberculosis  
 XX  
 PS Example 3; Page 159; 323pp; English.  
 XX  
 XX This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.  
 XX  
 XX Sequence 95 AA;  
 SQ

Query Match 100.0%; Score 137; DB 20; Length 95;

Best Local Similarity 100.0%; Pred. No. 2.4e-14;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 STNIRQAGVQYSRADEEQQALSSOMGF 28  
          |||||  
Db 68 stnirqagvqysradeeqqalssomgf 95

Search completed: July 5, 2001, 11:45:44  
Job time: 273 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:47:01 ; Search time 61.79 seconds  
(without alignments)  
9.128 Million cell updates/sec

Title: US-09-462-480-12  
Perfect score: 137  
Sequence: 1 STNIRQAGVQVSRADERQQALSSQMGF 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	31.4	60	US-08-463-081B-4	Sequence 4, Appli
2	43	31.4	60	US-08-461-379A-4	Sequence 4, Appli
3	43	31.4	60	US-08-462-390B-4	Sequence 4, Appli
4	43	31.4	60	US-08-463-074B-4	Sequence 4, Appli
5	43	31.4	60	US-08-465-585C-4	Sequence 4, Appli
6	43	31.4	60	US-08-652-446-4	Sequence 4, Appli
7	42	30.7	40	US-08-144-121-10	Sequence 10, Appl
8	42	30.7	40	US-08-735-893-10	Sequence 10, Appl
9	42	30.7	1147	US-08-144-121-3	Sequence 3, Appli
10	42	30.7	1147	US-08-735-893-3	Sequence 3, Appli
11	42	30.7	1165	US-08-144-121-2	Sequence 2, Appli
12	42	30.7	1165	US-08-735-893-2	Sequence 2, Appli
13	41.5	30.3	1080	US-08-162-081B-36	Sequence 36, Appl
14	41.5	30.3	1080	US-08-780-872-36	Sequence 36, Appl
15	41.5	30.3	1394	PCT-US95-10661A-2	Sequence 2, Appli
16	41.5	30.3	1477	US-08-492-459-10	Sequence 10, Appl
17	41.5	30.3	1477	US-08-423-752-10	Sequence 10, Appl
18	41.5	30.3	1477	US-08-945-994-3	Sequence 3, Appli
19	41.5	30.3	1477	US-08-716-873-24	Sequence 24, Appl
20	41.5	30.3	2414	US-08-227-536-2	Sequence 2, Appli
21	41.5	30.3	2414	PCT-US95-04682-2	Sequence 2, Appli
22	41	29.9	619	US-08-465-746-2	Sequence 2, Appli
23	41	29.9	619	US-08-214-164-2	Sequence 2, Appli
24	41	29.9	619	US-08-467-852A-3	Sequence 3, Appli
25	41	29.9	619	US-08-246-636-2	Sequence 2, Appli
26	41	29.9	619	US-08-247-491A-3	Sequence 3, Appli
27	41	29.9	619	US-08-319-795-2	Sequence 2, Appli

28	41	29.9	619	2	US-08-468-985-2	Sequence 2, Appli
29	41	29.9	619	3	US-08-312-949-2	Sequence 2, Appli
30	41	29.9	648	1	US-08-072-070-2	Sequence 2, Appli
31	41	29.9	648	1	US-08-469-434-2	Sequence 2, Appli
32	41	29.9	648	1	US-08-214-222-2	Sequence 2, Appli
33	41	29.9	648	2	US-08-467-852A-2	Sequence 2, Appli
34	41	29.9	648	2	US-08-468-718-2	Sequence 2, Appli
35	41	29.9	648	2	US-08-247-491A-2	Sequence 2, Appli
36	41	29.9	648	3	US-08-446-201-3	Sequence 3, Appli
37	41	29.9	695	1	US-08-127-499A-23	Sequence 23, Appl
38	41	29.9	695	1	US-08-482-847-23	Sequence 23, Appl
39	40.5	29.6	290	2	US-08-614-686A-1	Sequence 1, Appli
40	40	29.2	126	3	US-08-556-419-24	Sequence 24, Appl
41	40	29.2	366	2	US-08-928-692-28	Sequence 28, Appl
42	40	29.2	498	4	US-09-058-260-16	Sequence 16, Appl
43	40	29.2	503	2	US-08-781-802-2	Sequence 2, Appli
44	40	29.2	503	2	US-08-781-802-10	Sequence 10, Appl
45	40	29.2	503	2	US-08-781-802-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
US-08-463-081B-4  
; Sequence 4, Application US/08463081B  
; Patent No. 5871960  
; Patent No. 5871960 5837487  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Kendall A. & Beadling, Carol  
; TITLE OF INVENTION: Nucleic Acids Encoding CR5 Polypeptide,  
; VECTOR AND TRANSFORMED CELL THEREOF, AND EXPRESSION THEREOF  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PRETTY, SCHROEDER & POPLAWSKI  
; STREET: 444 South Flower St. - Suite 1900  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0,  
; SOFTWARE: Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,081B  
; FILING DATE: 5-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/104,736  
; FILING DATE: 10-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/796,066  
; FILING DATE: 20-NOV-91  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Viviana Amzel, Ph. D.  
; REGISTRATION NUMBER: 30,930  
; REFERENCE/DOCKET NUMBER: P66 38150 (DART-060)  
; TELEPHONE: (213) 622-7700  
; TELEFAX: (213) 489-4210  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 60 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-463-081B-4  
Query Match 31.4%; Score 43; DB 2; Length 60;  
Best Local Similarity 43.5%; Pred. No. 3.1;

Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Oy 5 ROAGVOYSRADEEQOQALSSOMG 27  
Db 23 RKAGVTLPKAEAEQOQSSGVCLG 45

## RESULT 2

US-08-461-379A-4  
; Sequence 4, Application US/08461379A  
; Patent No. 5871961  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Kendall A. & Beadling, Carol  
; TITLE OF INVENTION: Nucleic Acids Encoding CR5 Polypeptide,  
; TITLE OF INVENTION: Vector and Transformed Cell Thereof, and  
; TITLE OF INVENTION: Expression Thereof  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ratner & Prestia  
; CITY: Valley Forge  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19482

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0,  
SOFTWARE: Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,379A  
FILING DATE: 5-JUNE-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/330,108; 08/104,736  
FILING DATE: 27-OCT-1994; 10-AUG-1993 & 20-NOV-91  
ATTORNEY/AGENT INFORMATION:  
NAME: Viviana Amzel, Ph. D.  
REGISTRATION NUMBER: 30,930  
REFERENCE/DOCKET NUMBER: DART-070  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610)470-0700  
TELEFAX: (610)470-0701  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-461-379A-4

Query Match 31.4%; Score 43; DB 2; Length 60;  
Best Local Similarity 43.5%; Pred. No. 3.1;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Oy 5 ROAGVOYSRADEEQOQALSSOMG 27  
Db 23 RKAGVTLPKAEAEQOQSSGVCLG 45

## RESULT 3

US-08-462-390B-4  
; Sequence 4, Application US/08462390B  
; Patent No. 5882894  
; GENERAL INFORMATION:  
; APPLICANT: Smith, K. A., & Beadling, C.  
; TITLE OF INVENTION: Nucleic Acids Encoding CR8 Polypeptide, Vector and  
; TITLE OF INVENTION: Transformed Cell Thereof, and Expression Thereof  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ratner & Prestia

ADDRESSEE: (B) STREET:One Westlakes-Berwyn  
CITY: Valley Forge  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,390B  
FILING DATE: 5-JUNE-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/330,108  
FILING DATE: 27-OCT-1994  
APPLICATION NUMBER: USSN 08/104,736  
FILING DATE: 10-AUG-1993  
APPLICATION NUMBER: USSN 07/796,066  
FILING DATE: 20-NOV-91  
ATTORNEY/AGENT INFORMATION:  
NAME: Viviana Amzel, Ph. D.  
REGISTRATION NUMBER: 30,930  
REFERENCE/DOCKET NUMBER: DART-040  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610)407-0700  
TELEFAX: (610)407-0701  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-462-390B-4

Query Match 31.4%; Score 43; DB 2; Length 60;  
Best Local Similarity 43.5%; Pred. No. 3.1;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Oy 5 ROAGVOYSRADEEQOQALSSOMG 27  
Db 23 RKAGVTLPKAEAEQOQSSGVCLG 45

## RESULT 4

US-08-463-074B-4  
; Sequence 4, Application US/08463074B  
; Patent No. 6020155  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Kendall A. & Beadling, Carol  
; TITLE OF INVENTION: Nucleic Acids Encoding CR1 Fusion Protein, Vector an  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PRETTY, SCHROEDER & POPLAWSKI  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071

444 South Flower St. - Suite 1

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0,  
SOFTWARE: Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,074B  
FILING DATE: 5-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/104,736  
FILING DATE: 10-AUG-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/796,066  
FILING DATE: 20-NOV-91  
ATTORNEY/AGENT INFORMATION:  
NAME: Viviana Amzel, Ph. D.  
REGISTRATION NUMBER: 30,930  
REFERENCE/DOCKET NUMBER: P66 38143 (DART-020)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 622-7700  
TELEFAX: (213) 489-4210  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-463-074B-4

Query Match 31.4%; Score 43; DB 3; Length 60;  
Best Local Similarity 43.5%; Pred. No. 3.1;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 5 ROAGVOYSRADERQQALSSQMG 27  
Db 23 RKAGVTLPKRAEQSSGVSLG 45

RESULT 5  
US-08-465-585C-4  
Sequence 4, Application US/08465585C  
Patent No. 6027914  
GENERAL INFORMATION:  
APPLICANT: Smith, K. A., & Beadling, C.  
TITLE OF INVENTION: Nucleic Acids Encoding CR6 Polypeptide, Vector  
TITLE OF INVENTION: Transformed Cell Thereof, and Expression Thereof  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PRETTY, SCHROEDER & POPLAWSKI  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 900071  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 5-JUNE-1995  
APPLICATION NUMBER: US/08/465,585C  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/330,108  
FILING DATE: 27-OCT-1994  
APPLICATION NUMBER: USSN 08/104,736  
FILING DATE: 10-AUG-1993  
APPLICATION NUMBER: USSN 07/796,066  
FILING DATE: 20-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Viviana Amzel, Ph. D.  
REGISTRATION NUMBER: 30,930  
REFERENCE/DOCKET NUMBER: P66 38149 (DART-050)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 622-7700  
TELEFAX: (213) 489-4210  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-465-585C-4

444South Flower St. - Suite 190

Query Match 31.4%; Score 43; DB 3; Length 60;  
Best Local Similarity 43.5%; Pred. No. 3.1;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 5 ROAGVOYSRADERQQALSSQMG 27  
Db 23 RKAGVTLPKRAEQSSGVSLG 45

RESULT 6  
US-08-652-446-4  
Sequence 4, Application US/08652446  
Patent No. 6057427  
GENERAL INFORMATION:  
APPLICANT: Smith, Kendall A. & Beadling, Carol  
TITLE OF INVENTION: Nucleic Acids Encoding CR5  
TITLE OF INVENTION: Polypeptide, Vector and Transformed Cell Thereof, and  
TITLE OF INVENTION: Expression Thereof  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PRETTY, SCHROEDER & POPLAWSKI  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0,  
SOFTWARE: Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,446  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP App. # 96921319.8  
FILING DATE: 5-JAN-1998  
APPLICATION NUMBER: PCT/US/96/09194  
FILING DATE: 5-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/330,108  
FILING DATE: 27-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/463,074  
FILING DATE: 5-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/462,337  
FILING DATE: 5-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/462,390  
FILING DATE: 5-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/465,585  
FILING DATE: 5-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/463,081  
FILING DATE: 5-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/461,379  
FILING DATE: 5-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/739,523  
FILING DATE: 29-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Viviana Amzel, Ph. D.  
REGISTRATION NUMBER: 30,930  
REFERENCE/DOCKET NUMBER: FP66 40035  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 622-7700  
TELEFAX: (213) 489-4210  
INFORMATION FOR SEQ ID NO: 4:

444 South Flower St. - Suite 1

RESULT 8  
US-08-735-893-10  
; Sequence 10, Application US/08735893  
; Patent No. 5914317  
; GENERAL INFORMATION:

```

; ;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/144,121

```

;; FILING DATE: 27-OCT-1993  
;; CLASSIFICATION: 424  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: MYERS, Paul L.  
;; REGISTRATION NUMBER: 35,965  
;; REFERENCE/DOCKET NUMBER: (MGH-0780.0) MGP-021  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 227-7400  
;; TELEFAX: (617) 227-5941  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1147 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; FEATURE:  
;; NAME/KEY: Domain  
;; LOCATION: 1..231  
;; NAME/KEY: Domain  
;; LOCATION: 232..411  
;; FEATURE:  
;; NAME/KEY: Domain  
;; LOCATION: 412..765  
;; FEATURE:  
;; NAME/KEY: Domain  
;; LOCATION: 766..1147  
US-08-144-121-3

Query Match 30.7%; Score 42; DB 1; Length 1147;  
Best Local Similarity 39.3%; Pred. No. 1.7e+02;  
Matches 11; Conservative 5; Mismatches 8; Indels 4; Gaps 1;  
QY 5 RQAGVQVSRAD-EEQQALSSQMGF 28  
||| : : : :||| :||  
Db 1027 RQGAQVQVSRAD-EEQQALSSQMGF 1054

RESULT 10  
US-08-735-893-3  
; Sequence 3, Application US/08735893  
; Patent No. 5914317  
; GENERAL INFORMATION:  
; APPLICANT: Burgeson, Robert E.  
; APPLICANT: Wagon, David W.  
; TITLE OF INVENTION: BLK CHAIN OF LAMININ AND METHODS OF USE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: BOSTON  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/735,893  
; FILING DATE: 18-OCT-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/144,121  
; FILING DATE: 27-OCT-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MYERS, Paul L.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: (MGH-0780.1) MGP-021DV

;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 227-7400  
;; TELEFAX: (617) 227-5941  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1147 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; FRAGMENT TYPE: internal  
;; FEATURE:  
;; NAME/KEY: Domain  
;; LOCATION: 1..231  
;; NAME/KEY: Domain  
;; LOCATION: 232..411  
;; FEATURE:  
;; NAME/KEY: Domain  
;; LOCATION: 412..765  
;; FEATURE:  
;; NAME/KEY: Domain  
;; LOCATION: 766..1147  
US-08-735-893-3

Query Match 30.7%; Score 42; DB 2; Length 1147;  
Best Local Similarity 39.3%; Pred. No. 1.7e+02;  
Matches 11; Conservative 5; Mismatches 8; Indels 4; Gaps 1;  
QY 5 RQAGVQVSRAD-EEQQALSSQMGF 28  
||| : : : :||| :||  
Db 1027 RQGAQVQVSRAD-EEQQALSSQMGF 1054

RESULT 11  
US-08-144-121-2  
; Sequence 2, Application US/08144121  
; Patent No. 5610031  
; GENERAL INFORMATION:  
; APPLICANT: Burgeson, Robert E.  
; APPLICANT: Wagon, David W.  
; TITLE OF INVENTION: BLK CHAIN OF LAMININ AND METHODS OF USE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: BOSTON  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/144,121  
; FILING DATE: 27-OCT-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MYERS, Paul L.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: (MGH-0780.0) MGP-021  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1165 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-144-121-2

Query Match 30.7%; Score 42; DB 1; Length 1165;  
Best Local Similarity 39.3%; Pred. No. 1.7e+02;  
Matches 11; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

Qy 5 RQAGVQYSRADE---EQQQALSSQMGEF 28  
|||:|:|:|||||  
Db 1045 RQGAEAQVQAOQLAEGASEQALSQGEF 1072

## RESULT 12

US-08-735-893-2  
; Sequence 2, Application US/08735893  
; Patent No. 5914317  
; GENERAL INFORMATION:  
; APPLICANT: Burgeson, Robert E.  
; APPLICANT: Wagnan, David W.  
; TITLE OF INVENTION: BIK CHAIN OF LAMININ AND METHODS OF USE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: BOSTON  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/735,893  
; FILING DATE: 18-OCT-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/144,121  
; FILING DATE: 27-OCT-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Paul L.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: (MGH-0780.1) MGP-021DV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1165 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-735-893-2

Query Match 30.7%; Score 42; DB 2; Length 1165;  
Best Local Similarity 39.3%; Pred. No. 1.7e+02;  
Matches 11; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

Qy 5 RQAGVQYSRADE---EQQQALSSQMGEF 28  
|||:|:|:|||||  
Db 1045 RQGAEAQVQAOQLAEGASEQALSQGEF 1072

## RESULT 13

US-08-162-081B-36  
; Sequence 36, Application US/08162081B  
; Patent No. 5824492  
; GENERAL INFORMATION:  
; APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu  
; APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter  
; APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinia,

; APPLICANT: Stefano; Gout, Ivan Tarasovitch  
; TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/162,081B  
; FILING DATE: February 7, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB93/00761  
; FILING DATE: 13 April 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pasqualini, Patricia A.  
; REGISTRATION NUMBER: 34,894  
; REFERENCE/DOCKET NUMBER: LUD 5256  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1080 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-162-081B-36

Query Match 30.3%; Score 41.5; DB 2; Length 1080;  
Best Local Similarity 30.6%; Pred. No. 1.9e+02;  
Matches 11; Conservative 6; Mismatches 4; Indels 15; Gaps 1;

Qy 5 RQAGVQYSRA-----DEEQQALSSQ 25  
|||:|:|:|||||  
Db 502 REAGFSYSHAGLSNRLARNELRENDKQLKAISTR 537

RESULT 14  
US-08-780-872-36  
; Sequence 36, Application US/08780872  
; Patent No. 5846824  
; GENERAL INFORMATION:  
; APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu  
; APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter  
; APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinia,  
; APPLICANT: Stefano; Gout, Ivan Tarasovitch  
; TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB93/00761  
; FILING DATE: 13 April 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pasqualini, Patricia A.  
; REGISTRATION NUMBER: 34,894  
; REFERENCE/DOCKET NUMBER: LUD 5256  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 688-9200  
; TELEFAX: (212) 838-3884  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1080 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-162-081B-36

Query Match 30.3%; Score 41.5; DB 2; Length 1080;  
Best Local Similarity 30.6%; Pred. No. 1.9e+02;  
Matches 11; Conservative 6; Mismatches 4; Indels 15; Gaps 1;

Qy 5 RQAGVQYSRA-----DEEQQALSSQ 25  
|||:|:|:|||||  
Db 502 REAGFSYSHAGLSNRLARNELRENDKQLKAISTR 537



APPLICATION NUMBER: US/08/780,872  
FILING DATE: 09-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/162,081  
FILING DATE: February 7, 1994  
APPLICATION NUMBER: PCT/GB93/00761  
FILING DATE: 13 April 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pasqualini, Patricia A.  
REGISTRATION NUMBER: 34,894  
REFERENCE/DOCKET NUMBER: LUD 5256  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1080 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-780-872-36

Query Match 30.3%; Score 41.5; DB 2; Length 1080;  
Best Local Similarity 30.6%; Pred. No. 1.9e+02;  
Matches 11; Conservative 6; Mismatches 4; Indels 15; Gaps 1;

QY 5 RQGVQVYSRA-----DEEQQALSSQ 25  
|:|:| |:|:|  
Db 502 REAGFSYSHAGLSNRLARDNRENDKQLKAISTR 537  
|:|:| |:|:|

RESULT 15  
PCT-US95-10661A-2  
Sequence 2, Application PC/TUS9510661A  
GENERAL INFORMATION:  
APPLICANT: Washington University, et al.  
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10661A  
FILING DATE: 16-AUG-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,791  
FILING DATE: 25-AUG-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: EP-59941/RET  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1394 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
PCT-US95-10661A-2

Query Match 30.3%; Score 41.5; DB 5; Length 1394;  
Best Local Similarity 32.6%; Pred. No. 2.6e+02;  
Matches 15; Conservative 5; Mismatches 7; Indels 19; Gaps 3;

QY 2 TNTRQAGVQ-----YSRAD---EEQ---QQAALSSQMGF 28  
|:|:| |:|:| |:|:| |:|:|  
Db 1171 TNLRLQIGVQKALANGRIGAVFHSRSDNTFDEQVKNHATLTMMSGF 1216

Search completed: July 5, 2001, 11:47:02  
Job time: 316 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:48:36 ; Search time 79.63 Seconds  
(without alignments)  
26.785 Million cell updates/sec

Title: US-09-462-480-12  
Perfect score: 137  
Sequence: 1 STNIRQAGVQYSRADDEQQQALSSQMGF 28  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	100.0	100	2	H70802
2	57	41.6	100	2	T10032
3	55	40.1	743	2	D82883
4	52.5	38.3	1263	2	T15496
5	48	35.0	852	2	T08416
6	47	34.3	293	1	DAPSPC
7	46.5	33.9	291	2	D64043
8	46.5	33.9	1325	2	T42722
9	46	33.6	198	2	T13245
10	46	33.6	224	2	F82155
11	46	33.6	531	2	H72040
12	46	33.6	531	2	G86583
13	45	32.8	343	2	H82171
14	45	32.8	416	2	T10623
15	45	32.8	441	2	S76513
16	45	32.8	576	2	S63249
17	45	32.8	880	2	T38083
18	44.5	32.5	108	2	A70689
19	44	32.1	381	2	T34333
20	44	32.1	382	1	A60112
21	44	32.1	382	2	S15578
22	44	32.1	423	1	I40382
23	44	32.1	460	2	G70803
24	44	32.1	687	2	S19680
25	44	32.1	1209	2	H85839
26	44	32.1	1210	2	E64979
27	44	32.1	1961	1	A61231
28	44	32.1	1999	1	S21801
29	44	32.1	2160	2	T20241

30	43.5	31.8	80	2	D29674
31	43.5	31.8	253	2	F84258
32	43.5	31.8	478	2	T09896
33	43.5	31.8	1040	2	T25092
34	43	31.4	118	2	T32450
35	43	31.4	291	2	C71362
36	43	31.4	297	2	T20520
37	43	31.4	313	2	S07924
38	43	31.4	322	2	T04595
39	43	31.4	330	2	C83735
40	43	31.4	333	2	H71703
41	43	31.4	575	2	S17814
42	43	31.4	602	2	T02302
43	43	31.4	665	2	S75291
44	43	31.4	745	2	D96829
45	43	31.4	747	2	S71478

## ALIGNMENTS

RESULT 1

H70802

hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70802

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandram, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Sgares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A: Reference number: A70500; MUID: 98295987

A: Accession: H70802

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-100 <COL>

A: Cross-references: GB:AL022120; GB:AL123456; NID: g3261558; PIDN: CAA17966.1; PID: g296

A: Experimental source: strain H37RV

A: Genetics:

A: Gene: RV3874

Query Match 100.0%; Score 137; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 9.8e-14;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 STNIRQAGVQYSRADDEQQQALSSQMGF 28

|||||

Db 73 STNIRQAGVQYSRADDEQQQALSSQMGF 100

RESULT 2

T10032

hypothetical protein MLCB628.13c - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000

C:Accession: T10032

R: Eigmeier, K.; Honore, N.; Woods, S. A.; Caudron, B.; Cole, S. T.

Mol. Microbiol. 7, 197-206, 1993

A: Title: Use of an ordered cosmid library to deduce the genomic organization of Mycob

A: Reference number: Z16917; MUID: 93188700

A: Accession: T10032

A: Status: preliminary; translated from GB/EMBL/DDBJ

A: Molecule type: DNA

A: Residues: 1-100 <EIG>

A: Cross-references: EMBL: Y14967; NID: g2370268; PIDN: CAA75210.1; PID: g2370280

Query Match

Best Local Similarity 41.6%; Score 57; DB 2; Length 100;

Matches 10; Conservative 7; Mismatches 8; Indels 0; Gaps 0;



## RESULT 7

D64043  
citrate (pro-3S)-lyase (EC 4.1.3.6) beta chain - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 21-Jul-2000  
C:Accession: D64043  
R:Flerschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: D64043  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-291 <TIGR>  
A:Cross-references: GB:U32688; GB:L42023; NID:g1572966; PIDN:AAC21701.1; PID:g1572968; T  
C:Superfamily: Yersinia pestis hypothetical 29.7K protein  
C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 33.9%; Score 46.5; DB 2; Length 291;  
Best Local Similarity 39.4%; Pred. No. 18;  
Matches 13; Conservative 5; Mismatches 6; Indels 9; Gaps 2;

QY 5 ROAGVQ-----YSRADEEQ-----QQAETSSQMGF 28

Db 184 RAAGIQAFDTVYSNANNEGFLEKAAALIKQLGF 216

## RESULT 8

T42722  
male-enhanced antigen-2 - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000  
C:Accession: T42722  
R:Kondo, M.; Sutou, S.  
DNA Seq. 7, 71-82, 1997  
A:Title: Cloning and molecular characterization of cDNA encoding a mouse male-enhanced  
A:Reference number: Z22242; MUID:97217683  
A:Accession: T42722  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1325 <KON>  
A:Cross-references: EMBL:D78270; NID:d1096175; PID:d1020389; PIDN:BAAL19612.1  
A:Experimental source: strain CD-1  
C:Function:  
C:Keywords: leucine zipper

Query Match 33.9%; Score 46.5; DB 2; Length 1325;  
Best Local Similarity 38.5%; Pred. No. 1e+02;  
Matches 10; Conservative 10; Mismatches 5; Indels 1; Gaps 1;

QY 1 STNIRQAGVOYSRADEEQQAALSSQM 26

Db 280 STRI-QAQVEHSSQKQKQSLSEV 304

## RESULT 9

T13215  
hypothetical protein R198 - Lactobacillus phage phi-gle  
C:Species: Lactobacillus phage phi-gle  
C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000  
C:Accession: T13215  
R:Kodaira, K.I.; Oki, M.; Kakikawa, M.; Watanabe, N.; Hirakawa, M.; Yamada, K.; Taketo, G. J. Biol. Chem. 270, 45-53, 1997  
A:Title: Genome structure of the Lactobacillus temperate phage phi gle: the whole genome  
A:Reference number: Z17631; MUID:97225795  
A:Accession: T13215

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-198 <KOD>  
A:Cross-references: EMBL:X98106; NID:g1926320; PIDN:CAA66744.1; PID:g1926359  
C:Genetics:  
C:Note: Rorf198  
C:Superfamily: Lactobacillus phage phi-gle hypothetical protein R198

Query Match 33.6%; Score 46; DB 2; Length 198;  
Best Local Similarity 38.5%; Pred. No. 14;  
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 2 TNIRQAGVOYSRADEEQQAALSSQM 27

Db 169 TSLEAQDYVRLDDQDTQASLDSQM 194

## RESULT 10

F82150  
conserved hypothetical protein VC1853 [imported] - Vibrio cholerae (strain N16961 ser  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: F82150  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: F82150  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-224 <HEI>  
A:Cross-references: GB:AE004260; GB:AE003852; NID:g9556368; PIDN:AAF95001.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC1853  
A:Map position: 1

Query Match 33.6%; Score 46; DB 2; Length 224;  
Best Local Similarity 34.6%; Pred. No. 16;  
Matches 9; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 1 STNIRQAGVOYSRADEEQQAALSSQM 26

Db 195 SNLTLQQLWYSKASKEERAKFKWM 220

## RESULT 11

H72040  
conserved hypothetical protein Cpl126 [imported] - Chlamydomophila pneumoniae (strains  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 18-Aug-2000  
C:Accession: H72040; A81501  
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: H72040  
A:Molecule type: DNA  
A:Residues: 1-531 <ARN>  
A:Cross-references: GB:AE001656; GB:AE001363; NID:g4377047; PIDN:AAD18885.1; PID:g437  
A:Experimental source: strain CWL029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: A81501  
A:Molecule type: DNA  
A:Residues: 1-531 <REA>

Job time: 376 sec

Thu Jul 5 13:58:48 2001

us-09-462-480-12.rpr

Page 5

THIS PAGE BLANK (USPTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:51:45 ; Search time 41.8 Seconds  
(without alignments)  
22.946 Million cell updates/sec

Title: US-09-462-480-12  
Perfect score: 137  
Sequence: 1 STNIRQAGVYSRADEQQQALSSQMGF 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	47	34.3	292	1	BPHC_PSES1
2	46.5	33.9	291	1	CILB_HAEIN
3	46.5	33.9	1325	1	G160_MOUSE
4	46	33.6	2715	1	TRX2_HUMAN
5	45	32.8	576	1	YNI5_YEAST
6	45	32.8	634	1	SELB_MOOTH
7	45	32.8	880	1	TRK2_SCHPO
8	44	32.1	382	1	IPAC_SHIDY
9	44	32.1	382	1	IPAC_SHIFL
10	44	32.1	423	1	IDH_BACSU
11	44	32.1	687	1	TGLC_BOVIN
12	44	32.1	937	1	MSH2_NEUCR
13	44	32.1	1210	1	YEH1_ECOLI
14	44	32.1	1960	1	MYSN_HUMAN
15	43.5	31.8	80	1	PYS1_ANASP
16	43	31.4	313	1	GDA7_WHEAT
17	43	31.4	575	1	BCHE_RHOCA
18	43	31.4	634	1	SELB_DESBA
19	43	31.4	665	1	FTLB_SYNY3
20	43	31.4	745	1	HGL2_ARATH
21	43	31.4	826	1	VILI_CHICK
22	43	31.4	1119	1	ALS3_CANAL
23	43	31.4	2688	1	ZEP1_MOUSE
24	42.5	31.0	468	1	YQCL_CAEEL
25	42.5	31.0	945	1	RPN2_YEAST
26	42	30.7	255	1	YQIX_BACSU
27	42	30.7	436	1	RP54_BACSU
28	42	30.7	689	1	TGLC_CAVCU
29	42	30.7	1172	1	LMB3_HUMAN
30	41.5	30.3	80	1	PYS1_MASLA
31	41.5	30.3	302	1	CILB_ECOLI
32	41.5	30.3	1068	1	P11A_HUMAN
33	41.5	30.3	1394	1	HAP_HAEIN

34	41.5	30.3	1409	1	HAP1_HAEIN	P44596 haemophilus
35	41.5	30.3	2414	1	P300_HUMAN	O09472 homo sapien
36	41	29.9	201	1	IF3_MYCPN	P78024 mycoplasma
37	41	29.9	413	1	DXR_MYCTU	Q10798 mycobacteri
38	41	29.9	419	1	CBPA_BOVIN	P00730 bos taurus
39	41	29.9	481	1	THRC_CORGL	P23669 corynebacte
40	41	29.9	509	1	PR12_HUMAN	P49643 homo sapien
41	41	29.9	664	1	NTPI_ENTHR	P43439 enterococcu
42	41	29.9	732	1	YMM1_CAEEL	P34489 caenorhabdi
43	41	29.9	803	1	FPS_DRONE	P18106 drosophila
44	41	29.9	936	1	CAPP_RHOPA	O32483 rhodopseudo
45	41	29.9	1207	1	YL08_CAEEL	P34402 caenorhabdi

## ALIGNMENTS

RESULT 1	
BPHC_PSES1	
ID BPHC_PSES1 STANDARD; PRT; 292 AA.	
AC P17297; Q52441;	
DT 01-AUG-1990 (Rel. 15, Created)	
DT 01-NOV-1995 (Rel. 32, Last sequence update)	
DT 15-JUL-1999 (Rel. 38, Last annotation update)	
DE BIPHENYL-2,3-DIOL 1,2-DIOXYGENASE (EC 1.13.11.39) (23OHPB OXYGENASE)	
DE (2,3-DIHYDROXYBIPHENYL DIOXYGENASE) (DHBD).	
GN BPHC.	
OS Pseudomonas sp. (strain KKS102).	
OC Bacteria; Proteobacteria.	
OX NCBI_TaxID=307;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=89213965; PubMed=2540155;	
RA Kimbara K., Hashimoto T., Fukuda M., Koana T., Takegi M., Oishi M., Yano K.;	
RT "cloning and sequencing of two tandem genes involved in degradation of 2,3-dihydroxybiphenyl to benzoic acid in the polychlorinated biphenyl-degrading soil bacterium Pseudomonas sp. strain KKS102.";	
RL J. Bacteriol. 171:2740-2747(1989).	
RN [2]	
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).	
RX MEDLINE=96226036; PubMed=8636975;	
RA Senda T., Sugiyama K., Narita H., Yamamoto T., Kimbara K., Fukuda M., Sato M., Yano K., Mitsui Y.;	
RT "Three-dimensional structures of free form and two substrate complexes of an extradiol ring-cleavage type dioxygenase, the BphC enzyme from Pseudomonas sp. strain KKS102.";	
RL J. Mol. Biol. 255:735-752(1996).	
CC - - CATALYTIC ACTIVITY: BIPHENYL-2,3-DIOL + O(2) = 2-HYDROXY-6-OXO-6-PHENYLHEXA-2,4-DIENOATE + H(2)O.	
CC - - COFACTOR: FERROUS ION.	
CC - - PATHWAY: DEGRADATION OF BIPHENYLS AND POLYCHLOROBIPHENYLS (PCB) TO BENZOIC ACID AND CHLOROBENZOIC ACIDS.	
CC - - SUBUNIT: HOMOOCTAMER.	
CC - - SIMILARITY: BELONGS TO THE EXTRADIOL RING-CLEAVAGE DIOXYGENASE FAMILY.	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC -----	
CC EMBL; M26433; AAA25750.1; -.	
DR EMBL; D17319; BAA04141.1; -.	
DR PIR; A32312; DAPSPC.	
DR PDB; 1DHY; 15-OCT-95.	
DR InterPro; IPR000486; -.	
DR Pfam; PF01013; Extradiol_dioxy; 1.	
DR PROSITE; PS00082; EXTRADIOL_DIOXYGENAS; 1.	
KW Oxidoreductase; Dioxygenase; Aromatic hydrocarbons catabolism; Iron;	

OC Eukaryota; Metazoa; Chordata; Craniata; verte

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RA Angrand P.O., Valvatne H., Jeanmougin F., Adamson A.,  
RA van der Hoeven F., Olsen L., Teekotte H., Huang N., Poch O.,  
RA Lamerdin J., Chambon P., Losson R., Stewart A., Aasland R.,  
RT "Mammalian trithorax- and ASH1-like proteins: putative chromatin  
RT regulators which contain PHD fingers and SET domains";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RA Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,  
RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Ganes J., Danganan L.,  
RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,  
RA Carrano A.V.,  
RT "Sequence analysis of a 1 Mb region in human 19q13.1";  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).  
RX TISSUE=Brain;  
RX MEDLINE=97349984; PubMed=9205841;  
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,  
RA Wiedemann L.G., Chin S.-F., Muleris M., Batley S.J., Collins V.P.,  
RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps  
RT to 19q13.1 and is amplified in solid tumor cell lines";  
RL Oncogene 18:7975-7984(1999).  
RN [4]  
RP SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).  
RX TISSUE=Testis, and Leukocyte;  
RX MEDLINE=20105772; PubMed=10637508;  
RA Huntemann D.G., Chin S.-F., Muleris M., Batley S.J., Collins V.P.,  
RA Wiedemann L.G., Aparicio S., Caldas C.,  
RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps  
RT to 19q13.1 and is amplified in solid tumor cell lines";  
RL Oncogene 18:7975-7984(1999).  
RN [5]  
RP PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).  
RX TISSUE=Placenta, and Bone marrow;  
RX MEDLINE=99339983; PubMed=10409430;  
RA Fitzgerald K.T., Diaz M.O.,  
RT "MLL2: A new mammalian member of the trx/MLL family of genes";  
RL Genomics 59:187-192(1999).  
CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
CC TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.  
CC ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,  
CC SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL  
CC BLOOD LYMPHOCYTES, AND PLACENTA.  
CC -1- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.  
CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION  
CC FACTORS.  
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.  
CC -1- SIMILARITY: CONTAINS 3 PHD ZINC-FINGER DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ007041; CAB45385.1; -;  
DR EMBL; AD000671; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; AB002302; BAA20763.2; -;  
DR EMBL; AF186605; AAD56420.1; -;  
DR EMBL; AF104918; AAD17932.1; -;  
DR EMBL; AF105279; AAD26113.1; -;

DR EMBL; AF105280; AAD26112.1; -;  
DR InterPro: IPR001214; -;  
DR InterPro: IPR001965; -;  
DR InterPro: IPR002857; -;  
DR Pfam; PF00628; PHD; 3;  
DR Pfam; PF00856; SET; 1;  
DR Pfam; PF02008; zf-CXXC; 1;  
DR PROSITE; PS0280; SET; 1;  
KW DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;  
KW Transcription regulation; Alternative splicing;  
FT DNA\_BIND 37 44 A.T HOOK (BY SIMILARITY).  
FT DNA\_BIND 110 117 A.T HOOK (BY SIMILARITY).  
FT DNA\_BIND 357 365 A.T HOOK (BY SIMILARITY).  
FT ZN\_FING 1203 1252 PHD-TYPE 1.  
FT ZN\_FING 1253 1303 PHD-TYPE 2.  
FT ZN\_FING 1337 1396 PHD-TYPE 3.  
FT DOMAIN 1449 1471 BROMODOMAIN (DIVERGENT).  
FT DOMAIN 2586 2715 SET.  
FT DOMAIN 26 37 POLY-GLY.  
FT DOMAIN 248 255 POLY-PRO.  
FT DOMAIN 362 398 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 402 771 PRO-RICH.  
FT DOMAIN 808 812 POLY-GLN.  
FT DOMAIN 1963 1970 POLY-PRO.  
FT DOMAIN 2251 2259 POLY-PRO.  
FT VARSPPLIC 532 582  
FT VARSPLIC 583 582  
FT VARSPLIC 583 2715 MISSING (IN ISOFORM TRUNCATED).  
FT CONFLICT 834 834 K -> E (IN REF. 5).  
FT CONFLICT 941 941 S -> Y (IN REF. 5).  
FT CONFLICT 1317 1317 E -> Q (IN REF. 5).  
FT CONFLICT 1362 1362 H -> Y (IN REF. 5).  
FT CONFLICT 1438 1438 D -> N (IN REF. 5).  
FT CONFLICT 2622 2622 D -> H (IN REF. 5).  
SQ SEQUENCE 2715 AA; 293511 MW; C0615B981BBE7BF CRC64;

Query Match 33.6%; Score 46; DB 1; Length 2715;  
Best Local Similarity 34.8%; Pred. No. 1.5e+02;  
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 4 IRQAGVOYSDERQEQQAALSSQM 26  
::: || : ||: ||: ||:  
Db 796 LKRAKVQLFKIDQOOQKVAASM 818

RESULT 5  
ID YN15\_YEAR STANDARD; PRT; 576 AA.  
AC P53838;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE HYPOTHETICAL 65.0 KDA PROTEIN IN MET2-SEC2 INTERGENIC REGION.  
GN YNL275W OR N0626.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Messenguy F., Dubois E., Vierendeels F., Scherens B., Pierard A.,  
RA Glandsdorff N.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE ANION EXCHANGER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way





OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
RN NCBI\_TaxID=9913;  
RN  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 79-95; 157-166; 242-251 & 581-587.  
RC TISSUE=Artery;  
RX MEDLINE=92037637; PubMed=1682150;  
RT Nakaniishi K., Nara K., Hagiwara H., Aoyama Y., Ueno H., Hirose S.;  
RT "Cloning and sequence analysis of cDNA clones for bovine aortic-  
RT endothelial-cell transglutaminase.";  
RL Eur. J. Biochem. 202:15-21(1991).  
CC -!- FUNCTION: CATALYZES THE CROSS-LINKING OF PROTEINS AND THE  
CC CONJUGATION OF POLYAMINES TO PROTEINS.  
CC -!- CATALYTIC ACTIVITY: PROTEIN GLUTAMINE + ALKYLAMINE = PROTEIN  
CC N(5)-ALKYLGLUTAMINE + NH(3).  
CC -!- COFACTOR: CALCIUM IS REQUIRED FOR TRANSGLUTAMINASE ACTIVITY.  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS ARE DETECTED IN THE LUNG. LOWER  
CC LEVELS ARE FOUND IN THE LIVER, SPLEEN AND HEART, BUT NOT IN THE  
CC BRAIN.  
CC -!- INDUCTION: BY RETINOIC ACID.  
CC -!- SIMILARITY: BELONGS TO THE TRANSGLUTAMINASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collabora-  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no  
CC modified and this statement is not removed. Usage by and for commer-  
CC ential requires a license agreement (see http://www.isb-sib.ch/annou-  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; X60866; CAA43097.1; .  
DR HSP; P00488; 1FTE.  
DR InterPro; IPR001102; .  
DR InterPro; IPR002931; .  
DR Pfam; PF01841; Transglut\_core; 1.  
DR Pfam; PF00927; Transglutamin\_C; 1.  
DR Pfam; PF00868; Transglutamin\_N; 1.  
DR PROSITE; PS00547; TRANSGLUTAMINASES; 1.  
KW Transferase; Acyltransferase; Calcium-binding.  
KW ACT\_SITE 277 BY SIMILARITY.  
SQ SEQUENCE 687 AA; 77112 MW; 7BBA00F15E779944 CRC64;  
  
Query Match 32.1%; Score 44; DB 1; Length 687;  
Best Local Similarity 40.9%; Pred. No. 64;  
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps  
  
QY 7 AGVQYSRADEEQQAALSSQMGF 28  
| | :|||:: :|||  
Db 145 ADAVYLDSEEREQYVLTKQGF 166  
  
RESULT 12  
MSH2\_NEUCR  
ID MSH2\_NEUCR STANDARD; PRT; 937 AA.  
AC O13396.  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DNA MISMATCH REPAIR PROTEIN MSH2.  
GN MSH2.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
ON [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1VA;  
RA Huber D.H., Hausener G., Yamamoto K., Ishii C., Seidel-Rogol B.L.,  
RI Berstrand H.;  
RI Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.



```

CC -----
CC TO OTHER NONMUSCLE MYOSINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z82215; CAB05105.1; -
DR EMBL; M81105; AAA59888.1; -
DR EMBL; M69180; AAA61765.1; -
DR EMBL; M31013; AAA36349.1; -
DR HSSP; P08799; LMND.
DR MIM; 160775; -
DR InterPro; IPR000048; -
DR InterPro; IPR001609; -
DR InterPro; IPR002928; -
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Coiled coil; Actin-binding; Alkylation; ATP-binding;
KW Multigene family; Calmodulin-binding.
FT DOMAIN 1 836 GLOBULAR HEAD (S1).
FT DOMAIN 837 1960 ROD-LIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 837 1960 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP.
FT DOMAIN 654 676 ACTIN-BINDING.
FT DOMAIN 786 800 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 812 829 CALMODULIN-BINDING (BY SIMILARITY).
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 53 55 EAI -> RGH (IN REF. 3).
FT CONFLICT 660 660 T -> S (IN REF. 3).
FT CONFLICT 869 869 T -> M (IN REF. 4).
FT CONFLICT 931 931 C -> Y (IN REF. 4).
FT CONFLICT 1240 1241 KG -> GR (IN REF. 4).
FT CONFLICT 1350 1350 E -> EE (IN REF. 2).
FT CONFLICT 1764 1764 T -> A (IN REF. 2).
FT CONFLICT 1771 1771 S -> G (IN REF. 2).
SQ SEQUENCE 1960 AA; 226531 MW; 588F84BB8C10656F CRC64;

```

```

Query Match 32.1%; Score 44; DB 1; Length 1960;
Best Local Similarity 40.0%; Pred. No. 2.1e+02;
Matches 10; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 STNIRQAGVQYSRADEEQQLSSQ 25
   ||::|| | |::|| |::||
Db 1869 STRLQKLRQLEEEAEAAQANASR 1893

```

```

RESULT 15
PYSI_ANASP STANDARD; PRT; 80 AA.
AC P07124;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE PHYCOBILISOME 8.9 KDA LINKER POLYPEPTIDE, PHYCOCYANIN-ASSOCIATED, ROD
DE (L-8.9/R) (ROD CAPPING LINKER PROTEIN).
GN CPD.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87246520; PubMed=3109890;
RA Belknap W.R., Haselkorn R.;
RT "Cloning and light regulation of expression of the phycocyanin operon

```

```

RT of the cyanobacterium Anabaena. ";
RL EMBL J. 6:871-884(1987).
CC -----
CC -!- SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05239; CAA28865.1; -
DR PIR; D29674; D29674.
DR InterPro; IPR001685; -
DR Pfam; PF01383; CpcD; 1.
KW Phycobilisome; Photosynthesis.
SQ SEQUENCE 80 AA; 8895 MW; 52DF2D7DEF5444D5 CRC64;

```

```

Query Match 31.8%; Score 43.5; DB 1; Length 80;
Best Local Similarity 47.8%; Pred. No. 7;
Matches 11; Conservative 3; Mismatches 4; Indels 5; Gaps

```

```

QY 3 NTRQAG-----VQYSRADEEQQ 20
   ||::|| | |::|| |::||
Db 39 NTRSGSVFIVPYSRMNEYQR 61

```

```

Search completed: July 5, 2001, 11:51:46
Job time: 505 sec

```





THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:50:52 ; Search time 123.78 Seconds  
(without alignments)  
29.928 Million cell updates/sec

Title: US-09-462-480-12

Perfect score: 137

Sequence: 1 STNIRQAGVQYSRADEQQALSSQMGF 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_16:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_unclassified:\*

13: sp\_vertebrate:\*

14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	137	100.0	100	2	Q69739
2	57	41.6	100	2	Q33084
3	55	40.1	743	2	Q9PVY9
4	52.5	38.3	1263	5	Q09371
5	48	35.0	852	10	Q9SVK4
6	48	35.0	904	10	Q38834
7	47	34.3	210	5	Q9VUG3
8	47	34.3	293	2	P97039
9	47	34.3	293	2	Q9RBT1
10	47	34.3	1013	5	Q9VYT7
11	46.5	33.9	1447	11	Q9QYT3
12	46.5	33.9	1487	11	Q9QYT2
13	46	33.6	198	9	Q03936
14	46	33.6	224	2	Q9K024
15	46	33.6	404	13	Q9DDJ2
16	46	33.6	531	2	Q9ZTF9
17	46	33.6	1543	5	Q9VWA0
18	45.5	33.2	200	11	Q9EPJ3
19	45	32.8	343	2	Q9KRG7

20	45	32.8	416	10	Q9SVG6
21	45	32.8	441	2	Q55718
22	44.5	32.5	91	14	Q9DXN0
23	44.5	32.5	108	2	P71653
24	44	32.1	356	4	Q9NU84
25	44	32.1	381	5	Q09585
26	44	32.1	625	6	Q02717
27	44	32.1	925	5	Q9V7C5
28	44	32.1	1961	11	Q62812
29	44	32.1	1999	11	Q63731
30	44	32.1	2160	5	O17709
31	43.5	31.8	253	1	Q9HQ59
32	43.5	31.8	434	4	Q9H4F8
33	43.5	31.8	478	10	Q9STW1
34	43.5	31.8	1040	13	Q90840
35	43.5	31.8	2084	3	Q9HEC9
36	43	31.4	118	5	O17366
37	43	31.4	248	4	Q9NVV1
38	43	31.4	291	2	O83163
39	43	31.4	295	2	O50280
40	43	31.4	297	5	O62131
41	43	31.4	313	10	Q41529
42	43	31.4	313	10	Q41546
43	43	31.4	322	10	O65517
44	43	31.4	330	2	Q9KF14
45	43	31.4	333	2	Q9ZD93

#### ALIGNMENTS

RESULT 1

Q69739

ID Q69739

AC Q69739

DT 01-AUG-1998

DT 01-AUG-1998

DT 01-JUN-2000

DE HYPOTHETICAL 10.8 KDA PROTEIN.

GN LHP OR RV3874 OR MTV027.09

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,

RA Taylor K., Whitehead S., Barrrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

complete genome sequence.";

RL Nature 393:537-544 (1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RA Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;

RT early secreted antigenic target 6 kDa (ESAT-6).";

RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF004671; AAC83445.1;

DR EMBL; AF004671; AAC83445.1;

DR TubercuList; RV3874;

DR Hypothetical protein.

KW Hypothetical protein.

SQ SEQUENCE 100 AA; 10794 MW; 285F4FC96F55D194 CRC64;

SEQUENCE FROM N.A.  
RC STRAIN-SEROVAR 3;  
RP MEDLINE-20500219; PubMed-11048724;  
RX Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Cassell G.H.;  
RA "The complete sequence of the mucosal pathogen Ureaplasma  
RT urealyticum";  
RT Nature 407:757-762(2000).  
RL EMBL: AE002148; AAF30913.1; -.  
RR InterPro: IPR000212; -.  
DR Dfam: PF00580; UvrD-helicase; 1.  
SQ SEQUENCE 743 AA; 86740 MW; 9D412574673D71E7 CRC64;

RESULT	5
Q9SVK4	
ID	Q9SVK4 PRELIMINARY; PRT; 852 AA.
AC	Q9SVK4;
DT	01-MAY-2000 (TReMBLrel. 13, Created)
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT	01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE	POTATIVE DISEASE RESISTANCE PROTEIN.
FE	F18B3.230.
GN	Arabisopsis thaliana (Mouse-ear cress).
OS	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC	Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_Taxid=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Rieger M., Gabel C., Mueller-Auer S., Schaefer M., Zipp M.,
RA	Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C., Quetier F.,
RA	Salanoubat M.;
RL	Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RL	[2]
RP	SEQUENCE FROM N.A.
RP	

RA EU Arabidopsis sequencing project;  
 RL EMBL: AL049862; CAB42924.1;  
 DR InterPro: IPR000767;  
 DR InterPro: IPR002182;  
 DR Pfam: PF00931; NB-ARC; 1.  
 DR PRINTS: PR00364; DISEASERSIST.  
 SQ SEQUENCE 852 AA; 97039 MW; 4A1DE0EF393801B6 CRC64;

Query Match 35.0%; Score 48; DB 10; Length 852;  
 Best Local Similarity 34.8%; Pred. No. 45;  
 Matches 8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 TNIRQAGVOYSRADDEEQOQLSS 24

Db 684 TNLKRLGLSLTRGQDIEEELDS 706

RESULT 6

Q38834 PRELIMINARY; PRT; 904 AA.

AC Q38834;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE MYOSIN HEAVY CHAIN HOMOLOG (FRAGMENT).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots;  
 OC Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LANDSBERG ERECTA;

RA Rouse D.T., Hazlewood J.L.;

RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL: U19616; AAA63149.1;

DR EMBL: 13922; Arath:1426;13922.

DR InterPro: IPR000767;

DR InterPro: IPR002182;

DR Pfam: PF00931; NB-ARC; 1.

DR PRINTS: PR00364; DISEASERSIST.

FT NON\_TER 1

SQ SEQUENCE 904 AA; 103471 MW; 19699788D31F990E CRC64;

Query Match 35.0%; Score 48; DB 10; Length 904;

Best Local Similarity 34.8%; Pred. No. 48;

Matches 8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 TNIRQAGVOYSRADDEEQOQLSS 24

Db 736 TNLKRLGLSLTRGQDIEEELDS 758

RESULT 7

Q9VUG3

ID Q9VUG3 PRELIMINARY; PRT; 210 AA.

AC Q9VUG3;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DE CG5048 PROTEIN.

GN CG5048.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Eukaryota; Metazoa; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=107311132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003533; AAF49720.1;  
 DR FlyBase: FBgn0036437; CG5048.  
 SQ SEQUENCE 210 AA; 24080 MW; EF9673135F670C4E CRC64;

Query Match 34.3%; Score 47; DB 5; Length 210;

Best Local Similarity 29.2%; Pred. No. 14;

Matches 7; Conservative 10; Mismatches 7; Indels 0; Gaps

QY 3 NIRQAGVOYSRADDEEQOQLSSQM 26

Db 20 NORRGIDYSSDEESMVVNM 43

RESULT 8

P97099

ID P97099 PRELIMINARY; PRT; 293 AA.

AC P97099;

DT 01-MAY-1997 (TrEMBLrel. 03, Created)

DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE 2,3-DIHYDROXYBIPHENYL DIOXYGENASE.

GN BPHC.

OS Alkaligenes eutrophus (Ralstonia eutropha).

OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

OC Ralstonia.

OX NCBI\_TaxID=510;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A5; TRANSPOSON-TN4371;

RX MEDLINE=97188526; PubMed=9037111;

RA Merlin C., Springael D., Mergeay M., Toussaint A.;

RT "Organisation of the bph gene cluster of transposon Tn4371, encoding

enzymes for the degradation of biphenyl and 4-chlorobiphenyl

RT Compounds. ";  
 RL Mol. Genet. 253:499-506(1997).  
 DR EMBL: X97984; CAA66622.1; -;  
 DR HSSP: P17297; 1DHX.  
 DR InterPro: IPR000325; -;  
 DR InterPro: IPR000486; -;  
 DR Pfam: PF00903; Glyoxalase; 2.  
 DR ProDom: PD000977; -; 1.  
 DR PROSITE: PS00082; EXTRADIOL\_DIOXYGENASE; UNKNOWN\_1.  
 KW DIOXYGENASE.  
 SQ SEQUENCE 293 AA; 32341 MW; 85D1A8702450D61F CRC64;  
 Query Match 34.3%; Score 47; DB 2; Length 293;  
 Best Local Similarity 52.9%; Pred. No. 20;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 4 IROAGVOYSRADEEQQ 20  
 :|||||:|||||:  
 Db 80 LRQAGVAFTRGDEALMQ 96  
 RESULT 9  
 Q9RBT1 PRELIMINARY; PRT; 293 AA.  
 ID Q9RBT1  
 AC Q9RBT1  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE 2,3-DIHYDROXYBIPHENYL-1,2-DIOXYGENASE.  
 GN BPC2  
 OS Pseudomonas sp. SY5.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 ON NCBI\_TaxID=106145;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN-SY5;  
 RA Chung S.-Y., Na K.-S., Kubo M.;  
 RT "Pseudomonas sp. strain SY5, 2,3-dihydroxybiphenyl-1,2-dioxygenase,  
 bphC2.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF190706; AAF04140.1; -;  
 DR HSSP: P17297; 1DHX.  
 DR InterPro: IPR000325; -;  
 DR InterPro: IPR000486; -;  
 DR Pfam: PF00903; Glyoxalase; 2.  
 DR ProDom: PD000977; -; 1.  
 DR PROSITE: PS00082; EXTRADIOL\_DIOXYGENASE; UNKNOWN\_1.  
 KW DIOXYGENASE.  
 SQ SEQUENCE 293 AA; 32341 MW; 3C194CE9DEE8BD0 CRC64;  
 Query Match 34.3%; Score 47; DB 2; Length 293;  
 Best Local Similarity 52.9%; Pred. No. 20;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 4 IROAGVOYSRADEEQQ 20  
 :|||||:|||||:  
 Db 80 LRQAGVAFTRGDEALMQ 96  
 RESULT 10  
 Q9VYV7 PRELIMINARY; PRT; 1013 AA.  
 ID Q9VYV7  
 AC Q9VYV7  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE CG15740 PROTEIN.  
 GN CG15740.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 ON NCBI\_TaxID=10090;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Beeson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003487; AAF48101.1; -;  
 DR FlyBase: FBgn0030340; CG15740.  
 SQ SEQUENCE 1013 AA; 111594 MW; 02CFF5C1F9F7590 CRC64;  
 Query Match 34.3%; Score 47; DB 5; Length 1013;  
 Best Local Similarity 42.9%; Pred. No. 78;  
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
 QY 5 RQAGVOYSRADEEQQALSSQ 25  
 :|:|:|:|:|:|:|:  
 Db 577 QNADIVYGRSKSQQQALAPQ 597  
 RESULT 11  
 Q9QVT3 PRELIMINARY; PRT; 1447 AA.  
 ID Q9QVT3  
 AC Q9QVT3  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE MALE ENHANCED ANTIGEN 2/GOLGI AUTOANTIGEN, GOLGIN SUBFAMILY A, 3.  
 GN MEA2/GOLGA3  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 ON NCBI\_TaxID=10090;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Beeson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003487; AAF48101.1; -;  
 DR FlyBase: FBgn0030340; CG15740.  
 SQ SEQUENCE 1013 AA; 111594 MW; 02CFF5C1F9F7590 CRC64;

```

Query Match          33.9%; Score 46.5; DB 11; Length 1487;
Best Local Similarity 38.5%; Pred. No. 1.4e+02;
Matches 10; Conservative 10; Mismatches 5; Indels 1; Gaps 1;

QY      1 STNIRAGVOYSRADEEQQAALSSQM 26
        || : || |::| ::||::|::|::
Db       442 STRL-QAQUEHSHSQKQDLSSEV 466

RESULT 13
O03936 PRELIMINARY; PRT; 198 AA.
ID O03936 AC
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE LACTOBACILLUS BACTERIOPHAGE PHIGLE COMPLETE GENOMIC DNA..
GN RORF198..
OS Bacteriophage phigle.
OC Viruses.
OX NCBI_TaxID=52979;
RN [1]

```

RA	MEDLINE:9/223795; PubMed:90/3005;
RR	Kodaira K.I., Oki M., Kakikawa M., Watanabe N., Hirakawa M.,
RT	Yamada K., Taketo A.;
RT	"Genome structure of the Lactobacillus temperate phage phi g1e: the
RT	whole genome sequence and the putative promoter/repressor system.";
RL	Gene 187:45-53(1997).
RR	EMBL; X98106; CAA66744.1; -.
DR	SEQUENCE 198 AA: 22540
SO	SEQUENCE 198 AA: 22540 E2509A56CA0DE4B6 CRC64;

```

Query Match          33.6%; Score 46; DB 9; Length 198;
Best Local Similarity 38.5%; Pred. No.19;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY      2  TNRQAGVOYSRADEEQQOALSQMG 27
      |:::| | | | | | | | | | | | | |
DB      169  TSLLERADYRLDDQDTQASLDSQMG 194

RESULT 14
Q9KQZ4      Q9KQZ4      PRELIMINARY;      PRT;      224 AA.
ID AC Q9KQZ4;
AC Q9KQZ4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL PROTEIN VC1853.
GN VC1853.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
NCBI_TaxID=666;
RN [1]
SEQUENCE FROM N.A.
STRAIN=EL TOR N16961 / SEROTYPE O1;
MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RA cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004260; AAF95001.1; -.
DR TIGR; VC1853; -.
SO SEQUENCE 224 AA: 25135 MW: 685C69F83B8D2520 CRC64:

```

Search completed: July 5, 2001, 11:50:52  
Job time: 475 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:45:44 ; Search time 130.35 Seconds  
(without alignments)  
7.441 Million cell updates/sec

Title: US-09-462-480-13

Perfect score: 78

Sequence: 1 RADEQQQALSSQMGF 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_0601.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	100.0	16	AAV03713	M. tuberculosis LH
2	78	100.0	28	AAW32444	Mycobacterium tube
3	78	100.0	28	AAW81698	M. tuberculosis im
4	78	100.0	28	AAV39128	M. tuberculosis an
5	78	100.0	28	AAV03712	M. tuberculosis LH
6	78	100.0	42	AAV03707	M. tuberculosis LH
7	78	100.0	80	AAW32454	Mycobacterium tube
8	78	100.0	80	AAW32386	Mycobacterium tube
9	78	100.0	80	AAW81707	M. tuberculosis im
10	78	100.0	80	AAW64340	Mycobacterium tube
11	78	100.0	80	AAV39137	M. tuberculosis an

12	78	100.0	80	AAV38994	M. tuberculosis re
13	78	100.0	95	AAW32444	Mycobacterium tube
14	78	100.0	95	AAW32376	Mycobacterium tube
15	78	100.0	95	AAW81747	M. tuberculosis im
16	78	100.0	95	AAW64321	Mycobacterium tube
17	78	100.0	95	AAV32097	Mycobacterium tube
18	78	100.0	95	AAV39118	M. tuberculosis an
19	78	100.0	95	AAV38981	M. tuberculosis re
20	78	100.0	100	AAW81706	M. tuberculosis im
21	78	100.0	100	AAW64339	Mycobacterium tube
22	78	100.0	100	AAV39136	M. tuberculosis an
23	78	100.0	100	AAV38993	M. tuberculosis re
24	78	100.0	100	AAV03705	M. tuberculosis LH
25	78	100.0	100	AAV35218	M. tuberculosis RV3
26	78	100.0	100	AAV19845	Mycobacterium tube
27	78	100.0	802	AAW81746	M. tuberculosis fu
28	78	100.0	802	AAW64379	Mycobacterium anti
29	78	100.0	802	AAV32063	Mycobacterium tu
30	78	100.0	802	AAV39224	M. tuberculosis
31	78	100.0	802	AAV39176	M. tuberculosis fu
32	78	100.0	802	AAV39081	M. tuberculosis fu
33	78	100.0	802	AAV39033	M. tuberculosis fu
34	43	55.1	160	AAV32850	Eucalyptus grandis
35	40	51.3	687	AAV32074	Bovine transglut
36	40	51.3	691	AAV04711	Sequence of guinea
37	39	50.0	124	AAV13230	Arabidopsis thalia
38	39	50.0	714	AAV43552	Arabidopsis thalia
39	39	50.0	722	AAV43551	Arabidopsis thalia
40	39	50.0	726	AAV43550	Arabidopsis thalia
41	39	50.0	864	AAW04184	Murine interleukin
42	39	50.0	864	AAW61271	Mouse interleukin-
43	39	50.0	864	AAW92408	Murine IL-17R prot
44	39	50.0	864	AAV99935	Murine IL-17R prot
45	39	50.0	864	AAV97130	Murine interleukin

#### ALIGNMENTS

RESULT 1

AAV03713  
ID AAV03713 standard; Protein; 16 AA.

XX AC AAV03713;

XX XX 07-JUN-1999 (first entry)

XX DE M. tuberculosis LHP polypeptide antigenic fragment.

XX ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
KW immune response.

XX OS Mycobacterium tuberculosis.

XX PN WC9904005-AA1.

XX PD 28-JAN-1999.

XX PF 16-JUL-1998; 98WO-IB01091.

XX PR 16-JUL-1997; 97US-0052631.

XX PA (INSP ) INST PASTEUR.

XX PA (STAT-) STATENS SERUM INST.

XX PA Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX WPI; 1999-132249/11.

XX PT New nucleic acid containing regulator and LHP gene of Mycobacterium  
PT tuberculosis - useful in vaccines, for diagnosis, and for expression  
PT of heterologous proteins

XX Claim 21; Page 65; 88pp; English.

XX The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX Sequence 16 AA;

SQ

Query Match 100.0%; Score 78; DB 20; Length 16;

Best Local Similarity 100.0%; Pred. No. 5.5e-07;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RADEEQQALSSQMGF 16

Db 1 radeeqqalssqmgf 16

|||||

RESULT 2

AAW32460

ID AAW32460 standard; Protein; 28 AA.

XX

AC AAW32460;

XX

DT 09-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb38-1 Peptide 6.

XX

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9709428-A2.

XX

PD 13-MAR-1997.

XX

PF 30-AUG-1996; 96WO-US14674.

XX

PR 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 22-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659683.

XX

PA (CORI-) CORIXA CORP.

XX

PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedwick TH;

XX

DR WPI; 1997-192903/17.

XX

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are

PT useful in vaccines for prevention or treatment of tuberculosis, also

PT for diagnosis

XX

PS Example 3; Page 131; 168pp; English.

XX

CC A new immunogenic polypeptide has been developed comprising an

CC

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or

CC its variant differing only in conservative substitutions and/or

CC modifications). The present sequence represents a M.tuberculosis

CC antigen, Tb38-1 Peptide 6. The immunogenic protein, and fusion proteins

CC containing one or more of the proteins or one of the proteins plus

CC ESAT-6, are useful in vaccines, preferably when formulated with a

CC non-specific adjuvant, to induce an immune response against

CC M.tuberculosis (for treatment or prevention).

XX

SQ Sequence 28 AA;

Query Match 100.0%; Score 78; DB 18; Length 28;

Best Local Similarity 100.0%; Pred. No. 1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RADEEQQALSSQMGF 16

Db 13 radeeqqalssqmgf 28

|||||

RESULT 3

AAW81698

ID AAW81698 standard; Protein; 28 AA.

XX

AC AAW81698;

XX

DT 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide TB38-1 peptide 6.

XX

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;

KW vaccine; pharmaceutical; infection; diagnosis.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9816646-A2.

XX

PD 23-APR-1998.

XX

PF 07-OCT-1997; 97WO-US18293.

XX

PR 13-MAR-1997; 97US-0818112.

PR 11-OCT-1996; 96US-0730510.

XX

PA (CORI-) CORIXA CORP.

XX

PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;

XX

DR WPI; 1998-261042/23.

XX

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used

PT to develop products for the detection of M. tuberculosis infection

PT and for diagnosis, treatment and prevention of tuberculosis

XX

PS Disclosure; Page 123; 230pp; English.

XX

XX This sequence represents an immunogenic portion of a soluble

CC Mycobacterium tuberculosis (MT) antigen which can be used in a method

CC for inducing protective immunity against tuberculosis (TB). This

CC sequence can be formulated into vaccines and/or pharmaceutical

CC compositions for immunising against M. tuberculosis infection or may

CC be used for the diagnosis of tuberculosis.

XX

SQ Sequence 28 AA;

Query Match 100.0%; Score 78; DB 19; Length 28;

Best Local Similarity 100.0%; Pred. No. 1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RADEEQQALSSQMGF 16

Db 13 raeeeqqalssqmgf 28  
 |||

## RESULT 4

AAV39128  
 ID AAY39128 standard; peptide; 28 AA.

XX AC AAY39128;

XX DT 05-NOV-1999 (first entry)

XX DE M. tuberculosis antigen tb38-1 peptide 6 amino acid sequence.

XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test.

XX OS Mycobacterium tuberculosis.

XX PN WO9942076-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03268.

XX PR 05-MAY-1998; 98US-0072967.

XX PR 18-FEB-1998; 98US-0025197.

XX PA (CORI-) CORIYA CORP.

XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX DR WPI; 1999-527409/44.

XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
 XX skin tests and protective or therapeutic vaccines or compositions

XX PS Example 3; Page 118; 299pp; English.

XX CC The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AA21949 to AA219460 and AAY39083 to  
 CC AAY39225 are used in the exemplification of the present invention.

XX SQ Sequence 28 AA;

Query Match 100.0%; Score 78; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. No. 1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RADEEQQALSSQMGF 16

Db 13 raeeeqqalssqmgf 28

## RESULT 5

AAV03712  
 ID AAY03712 standard; protein; 28 AA.

XX AC AAY03712;

XX DT 07-JUN-1999 (first entry)

DE M. tuberculosis LHP polypeptide antigenic fragment.

XX ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
 KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
 KW immune response.

XX OS Mycobacterium tuberculosis.

XX PN WO9904005-A1.

XX PD 28-JAN-1999.

XX PF 16-JUL-1998; 98WO-IB01091.

XX PR 16-JUL-1997; 97US-0052631.

XX PA (INSP ) INST PASTEUR.

XX PA (STAT-) STATENS SEROM INST.

XX PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX PI WPI; 1999-132249/11.

XX PT New nucleic acid containing regulator and LHP gene of Mycobacterium  
 XX tuberculosis - useful in vaccines, for diagnosis, and for expression  
 XX of heterologous proteins

XX PS Claim 21; Page 65; 88pp; English.

XX CC The present invention is directed to a polynucleotide carrying the  
 CC regulatory expression signals of the ESAT-6 protein as well as an open  
 CC reading frame coding for an antigenic protein LHP from Mycobacterium  
 CC tuberculosis. Host cells comprising the polynucleotide are used for the  
 CC recombinant expression of the protein. The recombinant polypeptide can  
 CC be used as immunogens and vaccines, to protect against bacteria of the  
 CC M. tuberculosis complex in humans or animals (the vaccines may include  
 CC other immunogenic proteins of the bacteria or their fragments,  
 CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection in the  
 CC detection of specific antibodies. The regulatory region present in the  
 CC polynucleotide may be used to express almost any heterologous protein in  
 CC mycobacteria, particularly as a fusion with polyhistidine. The two  
 CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to  
 CC provide a synergistic increase in ability to induce a protective immune  
 CC response. Sequences AAY03706-713 represent antigenic fragments of the  
 CC LHP polypeptide.

XX SQ Sequence 28 AA;

Query Match 100.0%; Score 78; DB 20; Length 28;

Best Local Similarity 100.0%; Pred. No. 1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RADEEQQALSSQMGF 16

Db 13 raeeeqqalssqmgf 28

## RESULT 6

AAV03707  
 ID AAY03707 standard; protein; 42 AA.

XX AC AAY03707;

XX DT 07-JUN-1999 (first entry)

DE M. tuberculosis LHP polypeptide antigenic fragment.

XX ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
 KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
 KW immune response.

XX OS Mycobacterium tuberculosis.

XX WO9904005-A1.  
PN 28-JAN-1999.  
XX  
XX  
XX 16-JUL-1998; 98WO-1B01091.  
XX  
XX 16-JUL-1997; 97US-0052631.  
XX  
XX (INSP ) INST PASTEUR.  
XX (STAT-) STATENS SERUM INST.  
XX  
XX Andersen P, Berthet F, Gicquel B, Rasmussen PB;  
XX WPI; 1999-132249/11.  
XX  
XX New nucleic acid containing regulator and LHP gene of Mycobacterium  
XX tuberculosis - useful in vaccines, for diagnosis, and for expression  
XX of heterologous proteins  
XX  
XX Claim 21; Page 64; 88pp: English.  
XX  
XX The present invention is directed to a polynucleotide carrying the  
XX regulatory expression signals of the ESAT-6 protein as well as an open  
XX reading frame coding for an antigenic protein LHP from Mycobacterium  
XX tuberculosis. Host cells comprising the polynucleotide are used for the  
XX recombinant expression of the protein. The recombinant polypeptide can  
XX be used as immunogens and vaccines, to protect against bacteria of the  
XX M. tuberculosis complex in humans or animals (the vaccines may include  
XX other immunogenic proteins of the bacteria or their fragments,  
XX specifically ESAT-6); and (b) for diagnosing tuberculosis infection by  
XX detection of specific antibodies. The regulatory region present in the  
XX polynucleotide may be used to express almost any heterologous protein in  
XX mycobacteria, particularly as a fusion with polyhistidine. The two  
XX proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to  
XX provide a synergistic increase in ability to induce a protective immune  
XX response. Sequences AY03706-713 represent antigenic fragments of the  
XX LHP polypeptide.  
XX  
XX SQ Sequence 42 AA;  
  
Query Match 100.0%; Score 78; DB 20; Length 42;  
Best Local Similarity 100.0%; Pred. No. 1.6e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 RADEEQQQALSSQMGF 16  
DB 27 radeeqqqalssqmgf 42  
  
RESULT 7  
AAW32454  
ID AAW32454 standard; Protein; 80 AA.  
AC AAW32454;  
XX  
XX 09-JAN-1998 (first entry)  
XX  
XX Mycobacterium tuberculosis antigen Tb38-1F3.  
XX  
XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis.  
XX  
XX Mycobacterium tuberculosis.  
XX  
XX WO9709428-A2.  
XX  
XX 13-MAR-1997.  
XX  
XX 30-AUG-1996; 96WO-US14674.  
XX  
XX 12-JUL-1996; 96US-0680374.

DR WPI; 1997-192904/17.  
 XX N-PSDB; AAT91460.  
 PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 PT - useful for diagnosis of M. tuberculosis infection  
 XX  
 XX Example 3; Page 163; 190pp; English.  
 XX  
 XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M. tuberculosis  
 CC antigen, Tb38-1F3. The immunogenic polypeptide can be used to diagnose  
 CC M. tuberculosis infection by forming complexes with specific  
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
 CC polypeptide can be used as diagnostic primers or probes and agents  
 CC that bind to the antigen, especially monoclonal antibodies or  
 CC equivalent polyclonal antibodies, are also used for diagnosis.  
 XX  
 XX Sequence 80 AA;  
 SQ  
 Query Match 100.0%; Score 78; DB 18; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RADEEQQALSSQMGF 16  
 Db |||||  
 65 radeeqqalssqmgf 80  
 RESULT 9  
 AAW81707  
 ID AAW81707 standard; Protein; 80 AA.  
 XX  
 AC AAW81707;  
 XX  
 DT 27-JAN-1999 (first entry)  
 XX  
 DE M. tuberculosis immunogenic polypeptide Tb38-1F3.  
 XX  
 KW Tuberculosis; Immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9816646-A2.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 07-OCT-1997; 97WO-US18293.  
 XX  
 PR 13-MAR-1997; 97US-0818112.  
 PR 11-OCT-1996; 96US-0730510.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 WPI; 1998-261042/23.  
 DR N-PSDB; AAV64509.  
 XX  
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX  
 XX Example 3b; Page 139-140; 230pp; English.  
 XX  
 XX This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical

CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.  
 XX  
 SQ Sequence 80 AA;  
 Query Match 100.0%; Score 78; DB 19; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RADEEQQALSSQMGF 16  
 Db |||||  
 65 radeeqqalssqmgf 80  
 RESULT 10  
 AAW64340  
 ID AAW64340 standard; Protein; 80 AA.  
 XX  
 AC AAW64340;  
 XX  
 DT 09-NOV-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen Tb38-1F3.  
 DE  
 KW Tuberculosis; infection; diagnosis; antigen; Tb38-1F3.  
 XX  
 OS Mycobacterium tuberculosis strain H37Rv.  
 XX  
 PN WO9816645-A2.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 07-OCT-1997; 97WO-US18214.  
 XX  
 PR 13-MAR-1997; 97US-0818111.  
 PR 11-OCT-1996; 96US-0729622.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 WPI; 1998-251292/22.  
 DR N-PSDB; AAV44400.  
 XX  
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and diagnosis of tuberculosis  
 XX  
 XX Example 3; Page 146; 250pp; English.  
 XX  
 XX This polypeptide comprises Mycobacterium tuberculosis antigen  
 CC Tb38-1F3. A DNA molecule (see AAV44400) coding for Tb38-1F3 was  
 CC isolated from a M. tuberculosis strain H37Rv genomic library. The  
 CC invention relates to compositions and methods for diagnosing  
 CC tuberculosis. It provides polypeptides (see AAW64291-W64379)  
 CC comprising an antigenic portion of a soluble M. tuberculosis  
 CC antigen, or an immunogenic portion of an M. tuberculosis antigen,  
 CC as well as DNA sequences encoding such polypeptides, recombinant  
 CC expression vectors and transformed or transfected host cells. Also  
 CC claimed are methods and diagnostic kits for detecting M.  
 CC tuberculosis infection in a patient using these polypeptides,  
 CC antibodies or oligonucleotide probes and primers, for the diagnosis  
 CC of tuberculosis.  
 XX  
 XX Sequence 80 AA;  
 Query Match 100.0%; Score 78; DB 19; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RADEEQQALSSQMGF 16  
 |||||  
 Db 65 radeeqqalssqmgf 80

## RESULT 11

AA39137  
 ID AAY39137 standard; Protein; 80 AA.

XX  
 AC AAY39137;  
 XX

DT 05-NOV-1999 (first entry)  
 XX

DE M. tuberculosis antigen Tb38-1f3 amino acid sequence.  
 XX

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test.  
 XX

OS Mycobacterium tuberculosis.  
 XX

XX WO9942076-A2.  
 PN

XX 26-AUG-1999.  
 PD

XX 17-FEB-1999; 99WO-US03268.  
 PF

XX 05-MAY-1998; 98US-0072967.  
 PR

XX 18-FEB-1998; 98US-0025197.  
 PR

XX (CORI-) CORIXA CORP.  
 PA

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 PI

XX WPI: 1999-527409/44.  
 DR

XX N-PSDB; AAZ19310.  
 DR

PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX  
 PS Example 3; Page 134-135; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein-containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to  
 CC AAY39225 are used in the exemplification of the present invention.  
 XX

SQ Sequence 80 AA;

Query Match 100.0%; Score 78; DB 20; Length 80;

Best Local Similarity 100.0%; Pred. No. 3.3e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RADEEQQALSSQMGF 16

|||||

Db 65 radeeqqalssqmgf 80

## RESULT 12

AA38994  
 ID AAY38994 standard; Protein; 80 AA.

XX  
 AC AAY38994;  
 XX

DT 05-NOV-1999 (first entry)  
 XX

DE M. tuberculosis recombinant antigen protein Tb38-1f3.  
 XX

KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity.  
 XX

OS Mycobacterium tuberculosis.  
 XX

PN WO9942118-A2.  
 PN

XX 26-AUG-1999.  
 PD

XX 17-FEB-1999; 99WO-US03265.  
 PF

XX 05-MAY-1998; 98US-0072596.  
 PR

XX 18-FEB-1998; 98US-0024753.  
 PR

XX (CORI-) CORIXA CORP.  
 PA

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 PI

XX WPI: 1999-527416/44.  
 DR

XX N-PSDB; AAZ19098.  
 DR

PT New polypeptide comprising antigenic portions of M. tuberculosis  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX  
 PS Example 3; Page 180; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.  
 XX

SQ Sequence 80 AA;

Query Match 100.0%; Score 78; DB 20; Length 80;

Best Local Similarity 100.0%; Pred. No. 3.3e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RADEEQQALSSQMGF 16

|||||

Db 65 radeeqqalssqmgf 80

## RESULT 13

AAW32444  
 ID AAW32444 standard; Protein; 95 AA.

XX  
 AC AAW32444;  
 XX

DT 09-JAN-1998 (first entry)  
 DT

DE Mycobacterium tuberculosis antigen Tb38-1.  
 DE

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M. tuberculosis.  
 KW  
 OS Mycobacterium tuberculosis.  
 OS

XX WO9709428-A2.  
 PN

XX 13-MAR-1997.  
 PD

XX 30-AUG-1996; 96WO-US14674.  
 PF

XX 12-JUL-1996; 96US-0680574.  
 XX

XX 01-SEP-1995; 95US-0523436.  
 PR

PR 22-SEP-1995; 95US-0533634.  
 PR 22-MAR-1996; 96US-0620874.  
 PR 05-JUN-1996; 96US-0659683.  
 XX (CORI-) CORIXA CORP.  
 XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;  
 XX WPI; 1997-192903/17.  
 DR N-PSDB; AAT91509.  
 XX  
 XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
 PT useful in vaccines for prevention or treatment of tuberculosis, also  
 PT for diagnosis  
 XX  
 XX Example 3; Page 124; 168pp; English.  
 XX  
 XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M.tuberculosis  
 CC antigen, Tb38-1. The immunogenic protein, and fusion proteins  
 CC containing one or more of the proteins or one of the proteins plus  
 CC ESAT-6, are useful in vaccines, preferably when formulated with a  
 CC non-specific adjuvant, to induce an immune response against  
 CC M.tuberculosis (for treatment or prevention).  
 XX  
 XX Sequence 95 AA;

Query Match 100.0%; Score 78; DB 18; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RADEEQQALSSQMGF 16  
 Db 80 radeeqqalssqmgf 95  
 |||||

RESULT 14  
 AAW32376  
 ID AAW32376 standard; Protein; 95 AA.  
 XX  
 XX AC AAW32376;  
 XX  
 XX DT 13-JAN-1998 (first entry)  
 XX  
 XX DE Mycobacterium tuberculosis antigen Tb38-1.  
 XX  
 XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 XX skin testing; M.tuberculosis.  
 XX  
 XX OS Mycobacterium tuberculosis.  
 XX  
 XX PN W03709429-A2.  
 XX  
 XX PD 13-MAR-1997.  
 XX  
 XX PF 30-AUG-1996; 96WO-US14675.  
 XX  
 XX PR 12-JUL-1996; 96US-0680573.  
 XX 01-SEP-1995; 95US-0523435.  
 XX 22-SEP-1995; 95US-0532136.  
 XX 22-MAR-1996; 96US-0620280.  
 XX 05-JUN-1996; 96US-0658800.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;  
 XX WPI; 1997-192904/17.  
 DR

DR N-PSDB; AAT91445.  
 XX  
 XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 PT - useful for diagnosis of M. tuberculosis infection  
 XX  
 XX Example 3; Page 136; 190pp; English.  
 XX  
 XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M.tuberculosis  
 CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose  
 CC M.tuberculosis infection by forming complexes with specific  
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
 CC polypeptide can be used as diagnostic primers or probes and agents  
 CC that bind to the antigen, especially monoclonal antibodies or  
 CC equivalent polyclonal antibodies, are also used for diagnosis.  
 XX  
 XX Sequence 95 AA;

Query Match 100.0%; Score 78; DB 18; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 4e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RADEEQQALSSQMGF 16  
 Db 80 radeeqqalssqmgf 95  
 |||||

RESULT 15  
 AAW81747  
 ID AAW81747 standard; Protein; 95 AA.  
 XX  
 XX AC AAW81747;  
 XX  
 XX DT 27-JAN-1999 (first entry)  
 XX  
 XX DE M. tuberculosis immunogenic polypeptide Tb38-1.  
 XX  
 XX KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 XX vaccine; pharmaceutical; infection; diagnosis.  
 XX  
 XX OS Mycobacterium tuberculosis.  
 XX  
 XX PN W09816646-A2.  
 XX  
 XX PD 23-APR-1998.  
 XX  
 XX PF 07-OCT-1997; 97WO-US18293.  
 XX  
 XX PR 13-MAR-1997; 97US-0818112.  
 XX 11-OCT-1996; 96US-0730510.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 XX WPI; 1998-261042/23.  
 XX N-PSDB; AAV64491.  
 XX  
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX  
 XX Example 3b; Page 117; 230pp; English.  
 XX  
 XX This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This sequence  
 CC can be formulated into vaccines and/or pharmaceutical compositions for  
 CC immunising against M. tuberculosis infection or may be used for the

CC diagnosis of tuberculosis.

XX  
SQ Sequence 95 AA;

Query Match 100.0%; Score 78; DB 19; Length 95;  
Best Local Similarity 100.0%; Pred. No. 4e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RADEEQOQALSSQMGF 16  
| | | | | | | | | | | | | | | | | |  
Db 80 radeeqqalssqmgf 95

Search completed: July 5, 2001, 11:45:44  
Job time: 273 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:45:43 ; Search time 130.35 Seconds  
(without alignments)  
7.441 Million cell updates/sec

Title: US-09-462-480-11  
Perfect score: 77  
Sequence: 1 QEAANKQKQELDEIST 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues  
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601:\*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	77	100.0	16	AAV03711	M. tuberculosis LH
2	77	100.0	42	AAV03707	M. tuberculosis LH
3	77	100.0	80	AAW32454	Mycobacterium tube
4	77	100.0	80	AAW32386	Mycobacterium tube
5	77	100.0	80	AAW81707	M. tuberculosis im
6	77	100.0	80	AAW64340	Mycobacterium tube
7	77	100.0	80	AAV39137	M. tuberculosis an
8	77	100.0	80	AAV38994	M. tuberculosis re
9	77	100.0	95	AAW32444	Mycobacterium tube
10	77	100.0	95	AAW32376	Mycobacterium tube
11	77	100.0	95	AAW81747	M. tuberculosis im

12	77	100.0	95	19	AAW64321	Mycobacterium tube
13	77	100.0	95	20	AAV32097	Mycobacterium tube
14	77	100.0	95	20	AAV39118	M. tuberculosis an
15	77	100.0	95	20	AAV38981	M. tuberculosis re
16	77	100.0	100	19	AAW81706	Mycobacterium im
17	77	100.0	100	19	AAW64339	Mycobacterium tube
18	77	100.0	100	20	AAV39136	M. tuberculosis an
19	77	100.0	100	20	AAV38993	M. tuberculosis re
20	77	100.0	100	20	AAV03705	M. tuberculosis LH
21	77	100.0	100	22	AAW35218	M. tuberculosis RV3
22	77	100.0	100	22	AAW19845	Mycobacterium tube
23	77	100.0	802	19	AAW81746	M. tuberculosis fu
24	77	100.0	802	19	AAW64379	Mycobacterium anti
25	77	100.0	802	20	AAV32063	Mycobacterium tube
26	77	100.0	802	20	AAV39224	M. tuberculosis fu
27	77	100.0	802	20	AAV39176	M. tuberculosis fu
28	77	100.0	802	20	AAV39081	M. tuberculosis fu
29	77	100.0	802	20	AAV39033	M. tuberculosis fu
30	64	83.1	27	18	AAW32458	Mycobacterium tube
31	64	83.1	27	19	AAW81696	M. tuberculosis im
32	64	83.1	27	20	AAV39126	M. tuberculosis an
33	63	81.8	27	18	AAW32459	Mycobacterium tube
34	63	81.8	27	19	AAW81697	M. tuberculosis im
35	63	81.8	27	20	AAV39127	M. tuberculosis an
36	41	53.2	179	20	AAV50039	Proteus vulgaris t
37	41	53.2	344	21	AAW05574	Arabidopsis thalia
38	41	53.2	523	21	AAW05573	Arabidopsis thalia
39	41	53.2	539	21	AAW05572	Arabidopsis thalia
40	41	53.2	1411	17	AAW02258	Nucleolar/endosoma
41	40	51.9	181	21	AAW56895	Arabidopsis thalia
42	40	51.9	441	10	AAW90955	M6 streptococcal p
43	40	51.9	441	14	AAW41780	Streptococcus pyog
44	40	51.9	483	18	AAW08927	Type-6 M-protein.
45	39	50.6	103	16	AAW74928	Porcine elastin c

#### ALIGNMENTS

RESULT 1  
AAV03711  
ID AAV03711 standard; Protein; 16 AA.  
XX  
AC AAV03711;  
DT 07-JUN-1999 (first entry)  
XX  
DE M. tuberculosis LHP polypeptide antigenic fragment.  
XX  
ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
KW immune response.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO9904005-AL.  
XX  
PD 28-JAN-1999.  
XX  
PF 16-JUL-1998; 98WO-1B01091.  
XX  
PR 16-JUL-1997; 97US-0052631.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI (STAT-) STAFENS SERUM INST.  
XX  
PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;  
XX  
WPI; 1999-132249/11.  
XX  
New nucleic acid containing regulator and LHP gene of Mycobacterium  
PT tuberculosis - useful in vaccines, for diagnosis, and for expression  
PT of heterologous proteins

XX PS Claim 21; Page 65; 88pp; English.

CC The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX SQ Sequence 16 AA;

Query Match 100.0%; Score 77; DB 20; Length 16;

Best Local Similarity 100.0%; Pred. No. 8.4e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIST 16

Db 1 qeaankqkqeldeist 16

|||||

RESULT 2

AAY03707

ID AAY03707 standard; Protein; 42 AA.

XX AC AAY03707;

XX DT 07-JUN-1999 (first entry)

DE M. tuberculosis LHP polypeptide antigenic fragment.

XX ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;

KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;

KW immune response.

XX Mycobacterium tuberculosis.

OS WO9904005-A1.

PN 28-JAN-1999.

XX 16-JUL-1998; 98WO-IB01091.

PF 16-JUL-1997; 97US-0052631.

XX (INST ) INST PASTEUR.

PA (STAT-) STATENS SERUM INST.

XX Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX WPI; 1999-132249/11.

XX New nucleic acid containing regulator and LHP gene of Mycobacterium

PT tuberculosis - useful in vaccines, for diagnosis, and for expression

PT of heterologous proteins

XX Claim 21; Page 64; 88pp; English.

CC The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX SQ Sequence 42 AA;

Query Match 100.0%; Score 77; DB 20; Length 42;

Best Local Similarity 100.0%; Pred. No. 2.3e-05;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIST 16

Db 1 qeaankqkqeldeist 16

|||||

RESULT 3

AAW32454

ID AAW32454 standard; Protein; 80 AA.

XX AC AAW32454;

XX DT 09-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb38-1f3.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

OS WO9709428-A2.

PN 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.

PF 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 22-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659683.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

XX WPI; 1997-192903/17.

DR N-PSDB; AAT91526.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are

PT useful in vaccines for prevention or treatment of tuberculosis, also

PT for diagnosis

XX Example 3; Page 149; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or

CC its variant differing only in conservative substitutions and/or

CC modifications). The present sequence represents a M.tuberculosis

CC antigen, Tb38-1f3. The immunogenic protein, and fusion proteins

CC containing one or more of the proteins or one of the proteins plus

CC ESAT-6, are useful in vaccines, preferably when formulated with a

CC non-specific adjuvant, to induce an immune response against  
 CC M.tuberculosis (for treatment or prevention).

XX Sequence 80 AA;  
 SQ

Query Match 100.0%; Score 77; DB 18; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 16  
 |||||  
 Db 39 qeaankqkqeldeist 54

## RESULT 4

AAW32386  
 ID AAW32386 standard; Protein; 80 AA.

XX AC AAW32386;  
 XX AC

XX 13-JAN-1998 (first entry)  
 XX

DE Mycobacterium tuberculosis antigen Tb38-1F3.  
 XX

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M.tuberculosis.  
 XX

OS Mycobacterium tuberculosis.  
 XX

XX WO9709429-A2.  
 PN

XX 13-MAR-1997..  
 PD

XX 30-AUG-1996; 96WO-US14675.  
 PF

XX 12-JUL-1996; 96US-0680573.  
 PR

XX 01-SEP-1995; 95US-0523435.  
 PR

XX 22-SEP-1995; 95US-0532136.  
 PR

XX 22-MAR-1996; 96US-0620280.  
 PR

XX 05-JUN-1996; 96US-0658800.  
 PR

XX (CORI-) CORIXA CORP.  
 PA

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;

XX WPI; 1997-192904/17.  
 DR

XX N-PSDB; AAT91460.  
 DR

XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 PT - useful for diagnosis of M. tuberculosis infection

XX Example 3; Page 163; 190pp; English.  
 PS

XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M.tuberculosis  
 CC antigen, Tb38-1F3. The immunogenic polypeptide can be used to diagnose  
 CC M.tuberculosis infection by forming complexes with specific  
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
 CC polypeptide can be used as diagnostic primers or probes and agents  
 CC that bind to the antigen, especially monoclonal antibodies or  
 CC equivalent polyclonal antibodies, are also used for diagnosis.

XX Sequence 80 AA;  
 SQ

## Query Match

Best Local Similarity 100.0%; Score 77; DB 18; Length 80;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 16  
 |||||  
 Db 39 qeaankqkqeldeist 54

## RESULT 5

AAW81707  
 ID AAW81707 standard; Protein; 80 AA.

XX AC AAW81707;  
 XX AC

XX 27-JAN-1999 (first entry)  
 DT

XX M. tuberculosis immunogenic polypeptide Tb38-1F3.  
 DE

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 XX

OS Mycobacterium tuberculosis.  
 XX

XX WO9816646-A2.  
 PN

XX 23-APR-1998.  
 PD

XX 07-OCT-1997; 97WO-US18293.  
 PF

XX 13-MAR-1997; 97US-0818112.  
 PR

XX 11-OCT-1996; 96US-0730510.  
 PR

XX (CORI-) CORIXA CORP.  
 PA

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.  
 DR

XX N-PSDB; AAV64509.  
 DR

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX

PS Example 3B; Page 139-140; 230pp; English.  
 PS

XX This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.

XX Sequence 80 AA;  
 SQ

## Query Match

Best Local Similarity 100.0%; Score 77; DB 19; Length 80;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 16  
 |||||  
 Db 39 qeaankqkqeldeist 54

## RESULT 6

AAW64340  
 ID AAW64340 standard; Protein; 80 AA.

XX AC AAW64340;  
 XX AC

XX 09-NOV-1998 (first entry)  
 DT

XX Mycobacterium tuberculosis antigen Tb38-1F3.  
 DE

XX Tuberculosis; infection; diagnosis; antigen; Tb38-1F3.  
 KW

```

XX OS Mycobacterium tuberculosis strain H37Rv.
XX PN WO9816645-A2.
XX PD
XX PD 23-APR-1998.
XX PF
XX PF 07-OCT-1997; 97WO-US18214.
XX PR
XX PR 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX PA
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR
XX DR WPI; 1998-251292/22.
XX DR N-PSDB; AAV44400.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX PT and diagnosis of tuberculosis
XX PS
XX PS Example 3; Page 146; 250pp; English.
XX
XX This polypeptide comprises Mycobacterium tuberculosis antigen
XX CC Tb38-1F3. A DNA molecule (see AAV44400) coding for Tb38-1F3 was
XX CC isolated from a M. tuberculosis strain H37Rv genomic library. The
XX CC invention relates to compositions and methods for diagnosing
XX CC tuberculosis. It provides polypeptides (see AAV44291-W64379)
XX CC comprising an antigenic portion of a soluble M. tuberculosis
XX CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
XX CC as well as DNA sequences encoding such polypeptides, recombinant
XX CC expression vectors and transformed or transfected host cells. Also
XX CC claimed are methods and diagnostic kits for detecting M.
XX CC tuberculosis infection in a patient using these polypeptides,
XX CC antibodies or oligonucleotide probes and primers, for the diagnosis
XX CC of tuberculosis.
XX SQ Sequence 80 AA;

Query Match 100.0%; Score 77; DB 19; Length 80;
Best Local Similarity 100.0%; Pred. No. 4.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEAANKOKQELDEIST 16
Db | | | | | | | | | | | | | | | |
39 qeaankqkqeldeist 54

RESULT 7
AAV39137
ID AAV39137 standard; Protein; 80 AA.
XX
XX AC AAV39137;
XX
XX DT 05-NOV-1999 (first entry)
XX
XX DE M. tuberculosis antigen Tb38-1F3 amino acid sequence.
XX
XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;
XX KW immune response; skin test.
XX
XX OS Mycobacterium tuberculosis.
XX
XX PN WO9942076-A2.
XX
XX PD 26-AUG-1999.
XX
XX PF 17-FEB-1999; 99WO-US03268.

```

```

XX
XX 05-MAY-1998; 98US-0072967.
XX PR 18-FEB-1998; 98US-0025197.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX DR WPI; 1999-527409/44.
XX DR N-PSDB; AAZ19310.
XX
XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic
XX PT skin tests and protective or therapeutic vaccines or compositions
XX PS
XX PS Example 3; Page 134-135; 299pp; English.
XX
XX The present invention describes polypeptides comprising an immunogenic
XX CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
XX CC are vaccines and fusion protein containing M. tuberculosis Ag's.
XX CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
XX CC other polypeptides fragments, can be used in pharmaceutical compositions
XX CC or vaccines to generate a protective or therapeutic immune response to
XX CC M. tuberculosis and as reagents in skin tests for diagnosis of
XX CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
XX CC by, T, B or natural killer cells and/or macrophages in
XX CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
XX CC AAY39225 are used in the exemplification of the present invention.
XX SQ Sequence 80 AA;

Query Match 100.0%; Score 77; DB 20; Length 80;
Best Local Similarity 100.0%; Pred. No. 4.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEAANKOKQELDEIST 16
Db | | | | | | | | | | | | | | | |
39 qeaankqkqeldeist 54

RESULT 8
AAV38994
ID AAV38994 standard; Protein; 80 AA.
XX
XX AC AAV38994;
XX
XX DT 05-NOV-1999 (first entry)
XX
XX DE M. tuberculosis recombinant antigen protein Tb38-1F3.
XX
XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;
XX KW vaccine; immunity.
XX
XX OS Mycobacterium tuberculosis.
XX
XX PN WO9942118-A2.
XX
XX PD 26-AUG-1999.
XX
XX PF 17-FEB-1999; 99WO-US03265.
XX
XX PR 05-MAY-1998; 98US-0072596.
XX PR 18-FEB-1998; 98US-0024753.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
XX DR WPI; 1999-527416/44.
XX DR N-PSDB; AAZ19098.
XX

```

PT New polypeptide comprising antigenic portions of *M. tuberculosis*  
 XX  
 PS Example 3; Page 180; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from *Mycobacterium tuberculosis*. The novel  
 CC polypeptides are useful for detecting *M. tuberculosis* infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC *M. tuberculosis* infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.

XX Sequence 80 AA;

Query Match 100.0%; Score 77; DB 20; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QEAANKQKQELDEIST 16  
 Db 39 qeaankqkqeldeist 54  
 |||||

RESULT 9  
 AAW32444  
 ID AAW32444 standard; Protein; 95 AA.

XX AC AAW32444;

XX 09-JAN-1998 (first entry)

XX *Mycobacterium tuberculosis* antigen Tb38-1.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; *M. tuberculosis*.

XX *Mycobacterium tuberculosis*.

XX WO9709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.

XX 12-JUL-1996; 96US-0680574.

XX 01-SEP-1995; 95US-0523436.

XX 22-SEP-1995; 95US-0533634.

XX 22-MAR-1996; 96US-0620874.

XX 05-JUN-1996; 96US-0659683.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

XX Twardzik DR, Vedvick TH;

XX WPI; 1997-192903/17.

XX N-PSDB; AAT91509.

XX New immunogenic polypeptide(s) from *Mycobacterium tuberculosis* - are  
 PT useful in vaccines for prevention or treatment of tuberculosis, also  
 PT for diagnosis

XX Example 3; Page 124; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble *Mycobacterium tuberculosis* antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a *M. tuberculosis*  
 CC antigen, Tb38-1. The immunogenic protein, and fusion proteins  
 CC containing one or more of the proteins or one of the proteins plus  
 CC ESAT-6, are useful in vaccines, preferably when formulated with a  
 CC non-specific adjuvant, to induce an immune response against

CC *M. tuberculosis* (for treatment or prevention).  
 XX  
 SQ Sequence 95 AA;

Query Match 100.0%; Score 77; DB 18; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QEAANKQKQELDEIST 16  
 Db 54 qeaankqkqeldeist 69  
 |||||

RESULT 10

AAW32376

ID AAW32376 standard; Protein; 95 AA.

XX AC AAW32376;

XX 13-JAN-1998 (first entry)

XX *Mycobacterium tuberculosis* antigen Tb38-1.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; *M. tuberculosis*.

XX *Mycobacterium tuberculosis*.

XX WO9709429-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14675.

XX 12-JUL-1996; 96US-0680573.

XX 01-SEP-1995; 95US-0523435.

XX 22-SEP-1995; 95US-0532136.

XX 22-MAR-1996; 96US-0620280.

XX 05-JUN-1996; 96US-0658800.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

XX Twardzik DR, Vedvick TH;

XX WPI; 1997-192904/17.

XX N-PSDB; AAT91445.

XX New immunogenic polypeptide(s) from soluble *M. tuberculosis* antigens  
 PT - useful for diagnosis of *M. tuberculosis* infection

XX Example 3; Page 136; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble *Mycobacterium tuberculosis* antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a *M. tuberculosis*  
 CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose  
 CC *M. tuberculosis* infection by forming complexes with specific  
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
 CC polypeptide can be used as diagnostic primers or probes and agents  
 CC that bind to the antigen, especially monoclonal antibodies or  
 CC equivalent polyclonal antibodies, are also used for diagnosis.

XX Sequence 95 AA;

Query Match 100.0%; Score 77; DB 18; Length 95;

Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 QEAANKQKQELDEIST 16

Db 54 qeaankqkqeldeist 69  
|||||

RESULT 11  
AAW81747  
ID AAW81747 standard; Protein: 95 AA.

AC AAW81747;

DT 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide Tb38-1.

KB Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis.

OS Mycobacterium tuberculosis.

PN WO9816646-A2.

PD 23-APR-1998.

PF 07-OCT-1997; 97WO-US18293.

PR 13-MAR-1997; 97US-0818112.

PR 11-OCT-1996; 96US-0730510.

PA (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

DR WPI: 1998-261042/23.

DR N-PSDB; AAV64491.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PI to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis

PS Example 3b; Page 117; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble  
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
CC for inducing protective immunity against tuberculosis (TB). This sequence  
CC can be formulated into vaccines and/or pharmaceutical compositions for  
CC immunising against M. tuberculosis infection or may be used for the  
CC diagnosis of tuberculosis.

XX Sequence 95 AA;

Query Match 100.0%; Score 77; DB 19; Length 95;  
Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIST 16  
|||||

Db 54 qeaankqkqeldeist 69

RESULT 12  
AAW64321  
ID AAW64321 standard; Peptide: 95 AA.

AC AAW64321;

DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb38-1 peptide.

KB Tuberculosis; infection; diagnosis; antigen; Tb38-1.

XX

OS Mycobacterium tuberculosis strain H37Rv.

PN WO9816645-A2.

PD 23-APR-1998.

PF 07-OCT-1997; 97WO-US18214.

PR 13-MAR-1997; 97US-0818111.

PR 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

DR WPI: 1998-251292/22.

DR N-PSDB; AAV44384.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
PI to develop products for the detection of M. tuberculosis infection  
PT and diagnosis of tuberculosis

PS Example 3; Page 123; 250pp; English.

XX This is an antigenic portion of Mycobacterium tuberculosis antigen  
CC Tb38-1. A DNA sequence (see AAV44384) coding for antigen Tb38-1 was  
CC isolated from a M. tuberculosis strain H37Rv expression library  
CC using sera from patients having pulmonary or pleural tuberculosis.  
CC The invention relates to compositions and methods for diagnosing  
CC tuberculosis. It provides polypeptides (see AAV44384) comprising  
CC an antigenic portion of a soluble M. tuberculosis  
CC antigen, or an immunogenic portion of an M. tuberculosis antigen,  
CC as well as DNA sequences encoding such polypeptides, recombinant  
CC expression vectors and transformed or transfected host cells. Also  
CC claimed are methods and diagnostic kits for detecting M.  
CC tuberculosis infection in a patient using these polypeptides,  
CC antibodies or oligonucleotide probes and primers, for the diagnosis  
CC of tuberculosis.

XX Sequence 95 AA;

Query Match 100.0%; Score 77; DB 19; Length 95;  
Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIST 16  
|||||

Db 54 qeaankqkqeldeist 69

RESULT 13  
AAV32097  
ID AAV32097 standard; Protein: 95 AA.

AC AAV32097;

DT 17-JAN-2000 (first entry)

XX Mycobacterium tuberculosis antigen Tb38-1.

KB Tuberculosis; antigen; fusion protein; Tb38-1; diagnosis; therapy;  
KW vaccine; immunogen.

XX Mycobacterium tuberculosis.

XX WO9951748-A2.

PD 14-OCT-1999.

PF 07-APR-1999; 99WO-US07717.

XX

```

PR 07-APR-1998; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
PA (CORI-) CORIXA CORP.
PI Skelky YAW, Alderson M, Campos-Neto A;
XX WPI; 1999-601610/51.
XX New fusion proteins useful for diagnosis, prevention and treatment of
XX tuberculosis -
XX Claim 1; Fig 4D; 83pp; English.
XX This sequence represents the Mycobacterium tuberculosis antigen
XX Tb38-1. The invention provides fusion proteins (see AAY32059-71)
XX containing at least 2 M. tuberculosis antigens such as Tb38-1, e.g.
XX TbF-2 (see AAY32063) and a TbH9-Tb38-1 fusion. The new fusion
XX proteins are useful as vaccines for preventing tuberculosis
XX (claimed), for diagnosis (via in vitro assays or intradermal skin
XX tests for detection of anti-M. tuberculosis antibodies), monitoring
XX of disease progression, and treatment of tuberculosis. They are
XX more effective immunogens than mixtures of the individual protein
XX components.
XX
XX Sequence 95 AA;
SQ

```

Query Match 100.0%; Score 77; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 QEAANKQKQELDEIST 16
Db 54 qeaankqkqeldeist 69

```

RESULT 14  
 AAY39118  
 ID AAY39118 standard; Protein; 95 AA.  
 AC AAY39118;  
 XX  
 XX 05-NOV-1999 (first entry)  
 DT M. tuberculosis antigen Tb38-1 amino acid sequence.  
 DE  
 XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test.  
 XX  
 XX Mycobacterium tuberculosis.  
 OS  
 XX WO9942076-A2.  
 PN  
 XX 26-AUG-1999.  
 PD  
 XX 17-FEB-1999; 99WO-US03268.  
 PF  
 XX 05-MAY-1998; 98US-0072967.  
 PR 18-FEB-1998; 98US-0025197.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 PI WPI; 1999-527409/44.  
 XX  
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX  
 XX Example 3; Page 113; 299pp; English.

```

XX The present invention describes polypeptides comprising an immunogenic
XX part of a Mycobacterium tuberculosis antigen (Ag). Also described
XX are vaccines and fusion protein containing M. tuberculosis Ag's.
XX M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
XX other polypeptide fragments, can be used in pharmaceutical compositions
XX or vaccines to generate a protective or therapeutic immune response to
XX M. tuberculosis and as reagents in skin tests for diagnosis of
XX tuberculosis. Ag can induce proliferation of, or cytokine secretion
XX by, T, B or natural killer cells and/or macrophages in
XX tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
XX AAY39225 are used in the exemplification of the present invention.
XX
XX Sequence 95 AA;
SQ

```

Query Match 100.0%; Score 77; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 QEAANKQKQELDEIST 16
Db 54 qeaankqkqeldeist 69

```

RESULT 15  
 AAY38981  
 ID AAY38981 standard; Protein; 95 AA.  
 XX  
 AC AAY38981;  
 XX  
 XX 05-NOV-1999 (first entry)  
 DT M. tuberculosis recombinant antigen protein Tb38-1.  
 DE  
 XX Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity.  
 KW  
 XX Mycobacterium tuberculosis.  
 OS  
 XX WO9942118-A2.  
 PN  
 XX 26-AUG-1999.  
 PD  
 XX 17-FEB-1999; 99WO-US03265.  
 PF  
 XX 05-MAY-1998; 98US-0072596.  
 PR 18-FEB-1998; 98US-0024753.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 PI WPI; 1999-527416/44.  
 DR N-PSDB; AAZ19082.  
 DR  
 XX New polypeptide comprising antigenic portions of M. tuberculosis  
 PT  
 XX Example 3; Page 159; 323pp; English.  
 PS  
 XX This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.  
 CC  
 XX Sequence 95 AA;  
 SQ

Query Match 100.0%; Score 77; DB 20; Length 95;

Best Local Similarity 100.0%; Pred. No. 5.5e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 OEANKQKQELDEIST 16  
          |||||  
Db 54 qeankqkqeldest 69

Search completed: July 5, 2001, 11:45:44  
Job time: 273 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:47:01 ; Search time 61.79 Seconds  
(without alignments)  
5.216 Million cell updates/sec

Title: US-09-462-480-11  
Perfect score: 77  
Sequence: 1 QEAANKQKQELDE1ST 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	50.6	103	1	US-08-478-520-1
2	38	49.4	663	4	US-09-196-293-5
3	38	49.4	700	1	US-07-720-589-2
4	38	49.4	700	2	US-08-785-190-2
5	38	49.4	700	5	PCT-US92-05539-2
6	38	49.4	746	2	US-08-785-431-4
7	38	49.4	788	2	US-08-785-431-2
8	37	48.1	158	2	US-08-917-456-2
9	37	48.1	158	4	US-09-229-804-2
10	37	48.1	485	4	US-09-009-494-2
11	37	48.1	485	4	US-09-010-233-8
12	37	48.1	1394	5	PCT-US95-10661A-2
13	36	46.8	584	1	US-08-179-738-7
14	36	46.8	584	2	US-08-628-145-7
15	36	46.8	591	1	US-08-179-738-5
16	36	46.8	591	1	US-08-179-738-10
17	36	46.8	591	2	US-08-628-145-5
18	36	46.8	591	2	US-08-628-145-10
19	36	46.8	595	1	US-08-171-718-16
20	36	46.8	595	3	US-08-478-087-16
21	36	46.8	596	1	US-08-179-738-2
22	36	46.8	596	1	US-08-179-738-3
23	36	46.8	596	2	US-08-628-145-2
24	36	46.8	596	2	US-08-628-145-3
25	36	46.8	958	1	US-08-426-236-4
26	36	46.8	1162	2	US-08-728-323A-2
27	35	45.5	17	6	5304631-13

Patent No. 5304631

28 35 45.5 50 1 US-08-105-710-9 Sequence 9, Appli  
29 35 45.5 50 1 US-08-365-901-9 Sequence 9, Appli  
30 35 45.5 189 2 US-08-464-517-21 Sequence 21, Appl  
31 35 45.5 189 2 US-08-246-361A-21 Sequence 21, Appl  
32 35 45.5 189 3 PCT-US93-05000-21 Sequence 21, Appl  
33 35 45.5 189 5 PCT-US93-05000-21 Sequence 21, Appl  
34 35 45.5 289 2 US-08-246-361A-4 Sequence 4, Appli  
35 35 45.5 289 5 PCT-US93-05000-4 Sequence 4, Appli  
36 35 45.5 309 2 US-08-464-517-4 Sequence 4, Appli  
37 35 45.5 309 3 US-08-463-772-4 Sequence 4, Appli  
38 35 45.5 312 1 US-08-247-908A-2 Sequence 2, Appli  
39 35 45.5 312 1 US-08-453-942-2 Sequence 2, Appli  
40 35 45.5 312 2 US-08-926-885A-2 Sequence 2, Appli  
41 35 45.5 312 5 PCT-US94-05290-2 Sequence 2, Appli  
42 35 45.5 412 1 US-08-313-288B-18 Sequence 18, Appl  
43 35 45.5 423 2 US-08-760-797A-1 Sequence 1, Appli  
44 35 45.5 424 2 US-08-760-797A-3 Sequence 3, Appli  
45 35 45.5 424 4 US-08-932-929B-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-478-520-1  
; Sequence 1, Application US/08478520  
; Patent No. 5698519  
; GENERAL INFORMATION:  
; APPLICANT: Katsumura, No. 5698519uhiko  
; APPLICANT: Turk, Vito  
; TITLE OF INVENTION: POLYPEPTIDE SPECIFICALLY  
; TITLE OF INVENTION: INHIBITING CATHEPSIN L  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS  
; STREET: 2100 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,520  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP-283270/1993  
; FILING DATE: 12-NOV-1993  
; APPLICATION NUMBER: JP-313533/1993  
; FILING DATE: 14-DEC-1993  
; APPLICATION NUMBER: PCT/JP94/01878  
; FILING DATE: 08-NOV-1994  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 103 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-478-520-1

Query Match 50.6%; Score 39; DB 1; Length 103;  
Best Local Similarity 61.5%; Pred. No. 9.8;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 QEAANKQKQELDE 13  
|| ||| | : : :  
DB 22 QETANKVKFQLEE 34

RESULT 2  
US-09-196-293-5  
; Sequence 5, Application US/09196293  
; Patent No. 6183755  
; GENERAL INFORMATION:  
; APPLICANT: Fuchs, Renate  
; APPLICANT: Motz, Manfred  
; APPLICANT: Soutscheck, Erwin  
; APPLICANT: Wilske, Bettina  
; APPLICANT: Prec-Mursic, Vera  
; TITLE OF INVENTION: Active proteins from Borrelia  
; TITLE OF INVENTION: burgdorferi  
; FILE REFERENCE: 738.001US2  
; CURRENT APPLICATION NUMBER: US/09/196.293  
; CURRENT FILING DATE: 1998-11-19  
; EARLIER APPLICATION NUMBER: US 08/209,603  
; EARLIER FILING DATE: 1994-03-10  
; EARLIER APPLICATION NUMBER: US 07/862,535  
; EARLIER FILING DATE: 1992-06-19  
; EARLIER APPLICATION NUMBER: WO PCT/EP90/02282  
; EARLIER FILING DATE: 1990-12-21  
; EARLIER APPLICATION NUMBER: DE P39 42 728.5  
; EARLIER FILING DATE: 1989-12-22  
; EARLIER APPLICATION NUMBER: DE P40 18 988.0  
; EARLIER FILING DATE: 1990-06-13  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 663  
; TYPE: PRT  
; ORGANISM: Borrelia burgdorferi  
US-09-196-293-5

Query Match 49.4%; Score 38; DB 4; Length 663;  
Best Local Similarity 58.3%; Pred. No. 1e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EAANKQKQELDE 13  
:| |||:|  
Db 286 DAKKKQKEELDK 297

RESULT 3  
US-07-720-589-2  
; Sequence 2, Application US/07720589  
; Patent No. 5324630  
; GENERAL INFORMATION:  
; APPLICANT: Lefebvre, Rance B.  
; APPLICANT: Perrig, Guey-Chen  
; TITLE OF INVENTION: Methods and Compositions for Diagnosing  
; TITLE OF INVENTION: Lyme Disease  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: James M. Heslin  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/720,589  
; FILING DATE: 19910628  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 700 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-720-589-2

Query Match 49.4%; Score 38; DB 1; Length 700;  
Best Local Similarity 58.3%; Pred. No. 1.1e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EAANKQKQELDE 13  
:| |||:|  
Db 286 DAKKKQKEELDK 297

RESULT 4  
US-08-785-190-2  
; Sequence 2, Application US/08785190  
; Patent No. 5977339  
; GENERAL INFORMATION:  
; APPLICANT: Lefebvre, Rance B.  
; APPLICANT: Perrig, Guey-Chen  
; TITLE OF INVENTION: Methods and Compositions for Diagnosing  
; TITLE OF INVENTION: Lyme Disease  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: James M. Heslin  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,190  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/720,589  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heslin, James M.  
; REGISTRATION NUMBER: 29,541  
; REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 700 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-785-190-2

Query Match 49.4%; Score 38; DB 2; Length 700;  
Best Local Similarity 58.3%; Pred. No. 1.1e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EAANKQKQELDE 13  
:| |||:|  
Db 286 DAKKKQKEELDK 297

RESULT 5  
PCT-US92-05539-2  
Sequence 2, Application PC/TUS9205539  
GENERAL INFORMATION:  
APPLICANT: Lefebvre, Rance B.  
APPLICANT: Perng, Guey-Chen  
TITLE OF INVENTION: Methods and Compositions for Diagnosing  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: James M. Heslin  
STREET: 379 Lytton Avenue  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/04539  
FILING DATE: 19920629  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 23070-330/UC 90-017-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 700 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: Linear  
MOLECULE TYPE: protein  
PCT-US92-05539-2

Query Match 49.4%; Score 38; DB 5; Length 700;  
Best Local Similarity 58.3%; Pred. No. 1.1e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
QY 2 EAANKOKQELDE 13  
Db 286 DAKKKQKELDK 297

RESULT 6  
US-08-785-431-4  
Sequence 4, Application US/08785431  
Patent No. 5891667  
GENERAL INFORMATION:  
APPLICANT: Chalker, Allison  
APPLICANT: Hodgson, John  
TITLE OF INVENTION: No. 5891667el SpoIIIE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,431  
FILING DATE: 17-JAN-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9600955.0  
FILING DATE: 17-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31349  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 746 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-785-431-4

Query Match 49.4%; Score 38; DB 2; Length 746;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAANKOKQELDE 13  
Db 517 ELIRKQKQELDE 528

RESULT 7  
US-08-785-431-2  
Sequence 2, Application US/08785431  
Patent No. 5891667  
GENERAL INFORMATION:  
APPLICANT: Chalker, Allison  
APPLICANT: Hodgson, John  
TITLE OF INVENTION: No. 5891667el SpoIIIE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,431  
FILING DATE: 17-JAN-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9600955.0  
FILING DATE: 17-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31349  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 788 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-785-431-2

Query Match 49.4%; Score 38; DB 2; Length 788;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 EAANKQKQELDE 13  
| | | | |  
Db 559 ELIRKQKQELDE 570

RESULT 8  
US-08-917-456-2  
Sequence 2, Application US/08917456  
Patent No. 5866368  
GENERAL INFORMATION:  
APPLICANT: Greenwood, Rebecca C.  
APPLICANT: Gentry, Daniel R.  
TITLE OF INVENTION: NOVEL greA  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/917.456  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: GM10046  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 158 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-917-456-2

Query Match 48.1%; Score 37; DB 2; Length 158;  
Best Local Similarity 43.8%; Pred. No. 33;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIST 16  
| | | | |  
Db 11 QESFEKLELEELKT 26

RESULT 9  
US-09-229-804-2  
Sequence 2, Application US/09229804

Patent No. 6210674  
GENERAL INFORMATION:  
APPLICANT: Greenwood, Rebecca C.  
APPLICANT: Gentry, Daniel R.  
TITLE OF INVENTION: NOVEL greA  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: US  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/229,804  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/917,456  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dickinson, Todd Q  
REGISTRATION NUMBER: 28,354  
REFERENCE/DOCKET NUMBER: GM10046  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2252  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 158 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-229-804-2

Query Match 48.1%; Score 37; DB 4; Length 158;  
Best Local Similarity 43.8%; Pred. No. 33;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIST 16  
| | | | |  
Db 11 QESFEKLELEELKT 26

RESULT 10  
US-09-009-494-2  
Sequence 2, Application US/09009494  
Patent No. 6150340  
GENERAL INFORMATION:  
APPLICANT: Black, Michael T.  
APPLICANT: Lawlor, Elizabeth J.  
APPLICANT: Lewis, Ceri J.  
TITLE OF INVENTION: No. 6150340el Compounds  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/09/009,494  
;/ FILING DATE:  
;/ CLASSIFICATION:  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 60/037,857  
;/ FILING DATE: 07-FEB-1997  
;/ APPLICATION NUMBER: 60/044,365  
;/ FILING DATE: 28-APR-1997  
;/ APPLICATION NUMBER: 60/044,366  
;/ FILING DATE: 28-APR-1997  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Falk, Stephen T  
;/ REGISTRATION NUMBER: 36,795  
;/ REFERENCE/DOCKET NUMBER: GM50028  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 215-994-2488  
;/ TELEFAX: 215-994-2222  
;/ TELEX:  
;/ INFORMATION FOR SEQ ID NO: 2:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 485 amino acids  
;/ TYPE: amino acid  
;/ STRANDEDNESS: single  
;/ TOPOLOGY: linear  
;/ US-09-009-494-2

Query Match 48.1%; Score 37; DB 4; Length 485;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEI 14  
Db 48 KENAIKKAQELDEL 61

RESULT 11  
US-09-010-233-8  
;/ Sequence 8, Application US/09010233  
;/ Patent No. 6200774  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Black, Michael T.  
;/ APPLICANT: Lawlor, Elizabeth J.  
;/ APPLICANT: Lewis, Ceri J.  
;/ TITLE OF INVENTION: No. 6200774el Compounds  
;/ NUMBER OF SEQUENCES: 8  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: Dechert, Price & Rhoads  
;/ STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
;/ CITY: Philadelphia  
;/ STATE: PA  
;/ COUNTRY: USA  
;/ ZIP: 19103-2793  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Diskette  
;/ COMPUTER: IBM Compatible  
;/ OPERATING SYSTEM: DOS  
;/ SOFTWARE: FastSeq for Windows Version 2.0  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/09/010,233  
;/ FILING DATE:  
;/ CLASSIFICATION:  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 60/037,857  
;/ FILING DATE: 07-FEB-1997  
;/ APPLICATION NUMBER: 60/044,365  
;/ FILING DATE: 28-APR-1997  
;/ APPLICATION NUMBER: 60/044,366  
;/ FILING DATE: 28-APR-1997  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Falk, Stephen T  
;/ REGISTRATION NUMBER: 36,795

;/ REFERENCE/DOCKET NUMBER: GM50027  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 215-994-2488  
;/ TELEFAX: 215-994-2222  
;/ TELEX:  
;/ INFORMATION FOR SEQ ID NO: 8:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 485 amino acids  
;/ TYPE: amino acid  
;/ STRANDEDNESS: single  
;/ TOPOLOGY: linear  
;/ US-09-010-233-8

Query Match 48.1%; Score 37; DB 4; Length 485;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEI 14  
Db 48 KENAIKKAQELDEL 61

RESULT 12  
PCT-US95-10661A-2  
;/ Sequence 2, Application PC/TUS9510661A  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Washington University, et al.  
;/ TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
;/ NUMBER OF SEQUENCES: 9  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
;/ STREET: 4 Embarcadero Center, Suite 3400  
;/ CITY: San Francisco  
;/ STATE: California  
;/ COUNTRY: United States  
;/ ZIP: 94111-4187  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: Patent In Release #1.0, Version #1.25  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: PCT/US95/10661A  
;/ FILING DATE: 16-AUG-1995  
;/ CLASSIFICATION:  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: US 08/296,791  
;/ FILING DATE: 25-AUG-1994  
;/ CLASSIFICATION:  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Treccartin, Richard F.  
;/ REGISTRATION NUMBER: 31,801  
;/ REFERENCE/DOCKET NUMBER: FP-59941/RFT  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: (415) 781-1989  
;/ TELEFAX: (415) 398-3249  
;/ TELEX: 910 277299  
;/ INFORMATION FOR SEQ ID NO: 2:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 1394 amino acids  
;/ TYPE: amino acid  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: protein  
;/ PCT-US95-10661A-2

Query Match 48.1%; Score 37; DB 5; Length 1394;  
Best Local Similarity 53.3%; Pred. No. 3.3e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIS 15  
||| |||:: |||

Db 1097 QEEAEKQKQKDLIS 1111

## RESULT 13

US-08-179-738-7  
; Sequence 7, Application US/08179738  
; Patent No. 5578462  
; GENERAL INFORMATION:  
; APPLICANT: Seizinger, Bernd R.  
; APPLICANT: Kley, Nikolai A.  
; APPLICANT: Bianchi, Albert B.  
; TITLE OF INVENTION: No. 5578462el NF2 Isoforms  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Reed & Robins  
; STREET: 635 Bryant Street  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: U.S.A  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/179,738  
; FILING DATE: 10-JAN-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Robins, Roberta L.  
; REGISTRATION NUMBER: 33,208  
; REFERENCE/DOCKET NUMBER: 5998-0017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 617-8999  
; TELEFAX: (415) 327-3231  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 584 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: murine  
US-08-179-738-7

Query Match 46.8%; Score 36; DB 1; Length 584;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 16  
:||: ||: ||: ||: ||:  
Db 462 REAERRAKQKLEIAT 477

## RESULT 14

US-08-628-145-7  
; Sequence 7, Application US/08628145  
; Patent No. 5872214  
; GENERAL INFORMATION:  
; APPLICANT: Seizinger, Bernd R.  
; APPLICANT: Kley, Nikolai A.  
; APPLICANT: Bianchi, Albert B.  
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Reed & Robins  
; STREET: 635 Bryant Street  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: U.S.A

ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628,145  
; FILING DATE: 04-APR-1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/179,738  
; FILING DATE: 10-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Robins, Roberta L.  
; REGISTRATION NUMBER: 33,208  
; REFERENCE/DOCKET NUMBER: 5998-0017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 617-8999  
; TELEFAX: (415) 327-3231  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 584 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: murine  
US-08-628-145-7

Query Match 46.8%; Score 36; DB 2; Length 584;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 16  
:||: ||: ||: ||: ||:  
Db 462 REAERRAKQKLEIAT 477

## RESULT 15

US-08-179-738-5  
; Sequence 5, Application US/08179738  
; Patent No. 5578462  
; GENERAL INFORMATION:  
; APPLICANT: Seizinger, Bernd R.  
; APPLICANT: Kley, Nikolai A.  
; APPLICANT: Bianchi, Albert B.  
; TITLE OF INVENTION: No. 5578462el NF2 Isoforms  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Reed & Robins  
; STREET: 635 Bryant Street  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: U.S.A  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/179,738  
; FILING DATE: 10-JAN-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Robins, Roberta L.  
; REGISTRATION NUMBER: 33,208  
; REFERENCE/DOCKET NUMBER: 5998-0017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 617-8999

TELEFAX: (415) 327-3231  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 591 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: N-terminal  
 ORIGINAL SOURCE:  
 ORGANISM: murine  
 US-08-179-738-5

Query Match 46.8%; Score 36; DB 1; Length 591;  
 Best Local Similarity 50.0%; Pred. No. 2e+02;  
 Matches 8; Conservative 4; Mismatches 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 16  
 :|| :|| :|| :||  
 DB 462 REAERRAKQKLEIAT 477

Search completed: July 5, 2001, 11:47:01  
 Job time: 315 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 5, 2001, 11:48:35 ; Search time 79.63 seconds  
(without alignments)  
15.306 Million cell updates/sec

Title: US-09-462-480-11

Perfect score: 77

Sequence: 1 QEAANKQKQELDEIST 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	100.0	100	2 H70802	hypothetical prote
2	45	58.4	727	2 T29795	hypothetical prote
3	45	58.4	2116	2 A26650	myosin heavy chain
4	43	55.8	1011	2 T50544	poly(A)+ RNA trans
5	43	55.8	1012	2 T52000	poly(A)+ RNA trans
6	42	54.5	127	1 RBYUN	ubiquinol--cytochr
7	41	53.2	211	2 T47976	hypothetical prote
8	41	53.2	523	2 E96576	unknown protein, 4
9	41	53.2	610	2 S67701	hypothetical prote
10	41	53.2	627	2 D71729	dnak-type molecula
11	41	53.2	746	2 T47237	myosin II heavy ch
12	41	53.2	924	2 S06117	myosin heavy chain
13	41	53.2	1033	2 S73693	MG328 homolog p01
14	41	53.2	1410	~1 A57013	early endosome ant
15	41	53.2	2007	1 B43402	myosin heavy chain
16	40	51.9	181	2 T45782	hypothetical prote
17	40	51.9	483	2 A26297	M6 protein - Strep
18	40	51.9	699	2 T01029	hypothetical prote
19	40	51.9	1528	2 T06694	hypothetical prote
20	40	51.9	1560	2 T09202	probable tail comp
21	40	51.9	1826	2 T09271	probable tail comp
22	40	51.9	1976	2 A59252	myosin heavy chain
23	39	50.6	103	2 S40455	leukocyte cysteine
24	39	50.6	150	2 S60838	M protein precurs
25	39	50.6	300	2 S67189	hypothetical prote
26	39	50.6	434	2 T49044	hypothetical prote
27	39	50.6	521	2 B83947	hypothetical prote
28	39	50.6	643	1 FOLJLK	gag polyprotein -
29	39	50.6	775	2 T21259	hypothetical prote

30 39 50.6 1206 2 T34021 protein kinase SK2  
31 39 50.6 1233 2 T14157 serine/threonine p  
32 39 50.6 1805 1 A64224 hypothetical prote  
33 39 50.6 1853 1 A46761 myosin heavy chain  
34 38 49.4 182 2 G69402 conserved hypothet  
35 38 49.4 239 2 A44343 promotes sex organ  
36 38 49.4 250 2 T16337 hypothetical prote  
37 38 49.4 263 2 G84083 ABC transporter (A  
38 38 49.4 271 2 A84643 probable WRKY-type  
39 38 49.4 305 2 S77291 hypothetical prote  
40 38 49.4 356 2 A82152 toIA protein VC183  
41 38 49.4 440 2 G84631 hypothetical prote  
42 38 49.4 446 2 B82282 exodeoxyribonuclea  
43 38 49.4 449 2 T39410 hypothetical prote  
44 38 49.4 503 2 T06522 trans-cinnamate 4-  
45 38 49.4 521 2 C82922 methionyl-tRNA syn

#### ALIGNMENTS

##### RESULT 1

H70802

hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70802

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID: 98295987

A:Accession: H70802

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-100 <COL>

A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL7966.1; PID:g296

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3874

Query Match 100.0%; Score 77; DB 2; Length 100;

Best Local Similarity 100.0%; Pred. No. 3.2e-05;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 QEAANKQKQELDEIST 16

|||||

DB 59 QEAANKQKQELDEIST 74

##### RESULT 2

T29795

hypothetical protein T14G11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T29795

R:Fulton, B.; Hawkins, J.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid T14G11.

A:Reference number: Z20687

A:Accession: T29795

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-727 <FUL>

A:Cross-references: EMBL:U41024; PIDN:AAA82345.1; CESP:T14G11.3

C:Genetics:

A:Gene: CESP:T14G11.3

A:introns: 39/3; 54/3; 84/2; 140/2; 210/1; 292/3; 332/3; 370/3; 445/3; 573/3; 562/3

Query Match 58.4%; Score 45; DB 2; Length 727;  
Best Local Similarity 60.0%; Pred. No. 25;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIS 15  
||| ||| :||| :  
Db 313 QETANKLSHOLDEIN 327

RESULT 3  
A26655  
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)  
N:Contains: myosin ATPase (BC 3.6.1.32)  
C:Species: Dictyostelium discoideum  
C:Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 02-Feb-2001  
C:Accession: A26655; A24728; S00250  
R:Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.  
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986  
A:Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoid  
A:Reference number: A26655; MUID:87092266  
A:Accession: A26655  
A:Molecule type: DNA  
A:Residues: 1-2116 <WAR>  
A:Cross-references: GB:M14628; GB:M1938; NID:gl67834; PIDN:AAA33227.1; PID:gl67835  
R:DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985  
A:Reference number: A24728; MUID:86016788  
A:Accession: A24728  
A:Molecule type: mRNA  
A:Residues: 2035-2116 <DEL>  
R:Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.  
FEBS Lett. 227, 71-75, 1988  
A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium  
A:Reference number: S00250; MUID:88112226  
A:Accession: S00250  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1734-1893 <WAG>  
C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos  
F:1-818/Domain: globular head <HEP>  
F:89-747/Domain: myosin motor domain homology <MMOT>  
F:179-186/Region: nucleotide-binding motif A (P-loop)  
F:819-2116/Domain: alpha-helical rod <ROD>

Query Match 58.4%; Score 45; DB 2; Length 2116;  
Best Local Similarity 69.2%; Pred. No. 74;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EAANKQKQELDEI 14  
||| ||| :||| :  
Db 1339 EAKNKESELEI 1351

RESULT 4  
T50344  
poly(A)+ RNA transport protein Ptr3p [imported] - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C:Accession: T50344; T39493  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.  
submitted to the EMBL Data Library, February 2000  
A:Reference number: Z25063  
A:Accession: T50344  
A:Molecule type: DNA  
A:Residues: 1-644 <WOO>  
A:Cross-references: EMBL:AL139314; PIDN:CAB75417.1; GSPDB:GN000067; SPDB:SPBC211.09  
A:Experimental source: strain 972h(-); cosmid c211  
R:Beck, A.; Reinhardt, R.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, December 1998  
A:Reference number: Z21859

A:Accession: T39493  
A:Molecule type: DNA  
A:Residues: 611-1011 <BEC>  
A:Cross-references: EMBL:AL034433; PIDN:CAA22354.1; GSPDB:GN000066; SPDB:SPBC1604.21c  
A:Experimental source: strain 972h-; cosmid c1604  
C:Genetics: <WOO1>  
A:Gene: SPDB:SPBC211.09  
A:Map position: 2  
C:Genetics: <BEC1>  
A:Gene: ptr3; SPDB:SPBC1604.21c  
A:Map position: 1  
C:Superfamily: ubiquitin-activating enzyme E1

Query Match 55.8%; Score 43; DB 2; Length 1011;  
Best Local Similarity 60.0%; Pred. No. 72;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIS 15  
: ||| ||| :  
Db 777 ETRANKQKQELKSTA 791

RESULT 5  
T52000  
poly(A)+ RNA transport protein Ptr3p - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 03-Nov-2000  
C:Accession: T52000  
R:Azad, A.K.; Tani, T.; Shiki, N.; Tsuneyoshi, S.; Urushiyama, S.; Ohshima, Y.  
Mol. Biol. Cell 8, 825-841, 1997  
A:Title: Isolation and molecular characterization of mRNA transport mutants in Schizo  
A:Reference number: Z25898  
A:Accession: T52000  
A>Status: preliminary; translated from GE/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1012 <AZA>  
A:Cross-references: EMBL:D87259; PIDN:BAA75198.1  
C:Genetics:  
A:Gene: ptr3  
A:Introns: 2/3; 24/2  
C:Superfamily: ubiquitin-activating enzyme E1

Query Match 55.8%; Score 43; DB 2; Length 1012;  
Best Local Similarity 60.0%; Pred. No. 72;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QEAANKQKQELDEIS 15  
: ||| ||| :  
Db 777 ETRANKQKQELKSTA 791

RESULT 6  
RBYUN  
ubiquitin-cytochrome-c reductase (EC 1.10.2.2) 14k protein - yeast (Saccharomyces ce  
N:Alternate names: complex III 14k protein; cytochrome b-c1 complex 14k protein; prot  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Aug-1985 #sequence\_revision 13-Mar-1997 #text\_change 05-Nov-1999  
C:Accession: S69584; A00121  
R:Dietrich, F.S.  
submitted to the EMBL Data Library, August 1995  
A:Description: The sequence of S. cerevisiae cosmids 8166, 9787, 9717, and lambda 307  
A:Reference number: S69553  
A:Accession: S69584  
A:Molecule type: DNA  
A:Residues: 1-127 <DIE>  
A:Cross-references: EMBL:U33057; NID:g927764; PIDN:AAB64968.1; PID:g927796; GSPDB:GN0  
R:De Haan, M.; Van Loon, A.P.G.M.; Kreike, J.; Vaessen, R.T.M.J.; Grivell, L.A.  
Eur. J. Biochem. 138, 169-177, 1984  
A:Title: The biosynthesis of the ubiquinol-cytochrome c reductase complex in yeast. D  
A:Reference number: A00121; MUID:84108379  
A:Accession: A00121

A:Molecule type: DNA  
A:Residues: 1-91,'Q',93-127 <DEH>  
A:Cross-references: EMBL:X00256; NID:g3600; PIDN:CAA25064.1; PID:g3601  
A:Experimental source: strain FL100  
C:Genetics:  
A:Gene: SGD:OCR7; CRO1; UCR7; COR4; MIPS:YDR529c  
A:Cross-references: SGD:S0002937; MIPS:YDR529c  
A:Map position: 4R  
A:Genome: nuclear  
C:Superfamily: ubiquinol--cytochrome-c reductase 14K protein  
C:Keywords: electron transfer; membrane protein; membrane-associated complex; mitochondria

Query Match 54.5%; Score 42; DB 1; Length 127;  
Best Local Similarity 69.2%; Pred. No. 13;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EAANKQKQELDEI 14  
Db 111 EAAAKEKQELDNI 123  
|||||

RESULT 7  
T47976  
hypothetical protein F21F14.10 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 05-May-2000  
C:Accession: T47976  
R:Cholme, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; S...  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24481  
A:Accession: T47976  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-211 <CHO>  
A:Cross-references: EMBL:AL138642  
A:Experimental source: cultivar Columbia; BAC clone F21F14  
C:Genetics:  
A:Map position: 3  
A:Introns: 195/1  
A:Note: F21F14.10  
C:Superfamily: Arabidopsis thaliana hypothetical protein F21F14.10

Query Match 53.2%; Score 41; DB 2; Length 211;  
Best Local Similarity 66.7%; Pred. No. 30;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 NKQKQELDEIST 16  
Db 28 NKQKQERDMVKT 39  
|||||

RESULT 8  
E96576  
unknown protein, 43598-45751 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E96576  
R:theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: E96576  
A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-523 <STO>  
A:Cross-references: GB:AE005173; NID:g10645365; PIDN:AG21485.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F22G10.8  
A:Map position: 1

Query Match 53.2%; Score 41; DB 2; Length 523;  
Best Local Similarity 61.5%; Pred. No. 76;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 ANKQKQELDEIST 16  
Db 463 AKKQNEELDRIAT 475  
|||||

RESULT 9  
S67701  
hypothetical protein YDL153c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D1545  
C:Species: Saccharomyces cerevisiae  
C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 20-Jun-2000  
C:Accession: S67701  
R:Perea, J.; Blugeon, C.; Delaveau, T.; Jacq, C.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67693  
A:Accession: S67701  
A:Molecule type: DNA  
A:Residues: 1-610 <PER>  
A:Cross-references: EMBL:Z74201; NID:g1431239; PIDN:CAA98726.1; PID:g1431240; MIPS:YD  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:SAS10  
A:Cross-references: SGD:S0002312; MIPS:YDL153c  
A:Map position: 4L.

Query Match 53.2%; Score 41; DB 2; Length 610;  
Best Local Similarity 66.7%; Pred. No. 89;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EAANKQKQELDE 13  
Db 342 EDGGKQKQEI 353  
|||||

RESULT 10  
D71729  
dnak-type molecular chaperone RPi85 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: D71729  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499  
A:Accession: D71729  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-627 <AND>  
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14651.1; PID:g386  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: dnak; RPi85  
C:Function:

A:Description: Involved in protein folding and assembling/disassembling of protein co  
A:Superfamily: heat shock protein 70  
C:Keywords: ATP; molecular chaperone

Query Match 53.2%; Score 41; DB 2; Length 627;  
Best Local Similarity 56.2%; Pred. No. 91;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;



A:Molecule type: mRNA  
A:Residues: 1-2007 <TA2>  
A:Cross-references: GB:M93676; NID:g212448; PID:AAA48988.1; PID:g212452  
A>Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide sequence  
A:Accession: A43402  
A:Molecule type: mRNA  
A:Residues: 1-211;222-631;653-2007 <TA2>  
A:Cross-references: GB:M93676; NID:g212448; PID:AAA48988.1; PID:g212449  
A>Note: sequence extracted from NCBI backbone (NCBIN:112864)  
C:Comment: Alternately spliced segments 1 and 2 are found exclusively in nonmuscle myosin  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated  
F:1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>  
F:1-211,222-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYN>  
F:88-802/Domain: myosin motor domain homology <MMOT>  
F:178-185/Region: nucleotide-binding motif A (P-loop)  
F:212-221/Region: alternatively spliced segment 1 #status experimental  
F:559-593/Region: actin binding #status predicted  
F:632-652/Region: alternatively spliced segment 2 #status experimental  
F:692-714/Region: actin binding #status predicted  
F:875-2007/Domain: coiled coil #status predicted <COI>  
F:875-1315/Region: S2  
F:1316-2007/Region: light meromyosin  
F:129/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
F:184/Binding site: ATP (Lys) #status predicted  
F:732,742/Active site: Cys #status predicted  
F:1954/Binding site: phosphate (Thr) (covalent) #status predicted  
F:1987/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 53.2%; Score 41; DB 1; Length 2007;  
Best Local Similarity 53.3%; Pred. No. 3e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 2 EAANKQKQELDEIST 16  
DB 1311 EKANKLQNELDNVSS 1325

Search completed: July 5, 2001, 11:48:36  
Job time: 375 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:51:44 ; Search time 41.8 Seconds  
(without alignments)  
13.112 Million cell updates/sec

Title: US-09-462-480-11

Perfect score: 77

Sequence: 1 QEAANKQKQELDEIST 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	58.4	2116	1 MYS2_DICDI	P08799 dictyosteli
2	43	55.8	1012	1 UBAL_SCHPO	O94609 schizosacch
3	42	54.5	127	1 UCR_YEAST	P00128 saccharomyc
4	41	53.2	179	1 IF3_PROVU	P33319 proteus vul
5	41	53.2	627	1 DNAK_RICPR	Q9ZDX9 rickettsia
6	41	53.2	1033	1 Y328_MYCPN	P75310 mycoplasma
7	40	51.9	483	1 M6_STRPY	P08089 streptococ
8	40	51.9	1976	1 MYSO_HUMAN	P35580 homo sapien
9	39	50.6	35	1 CPI2_PIG	P80736 sus scrofa
10	39	50.6	103	1 CPI1_PIG	P35479 sus scrofa
11	39	50.6	103	1 CYT5_PIG	Q28986 sus scrofa
12	39	50.6	103	1 CYT8_PIG	Q28987 sus scrofa
13	39	50.6	643	1 GAG_SRF3L	P27400 simian foam
14	39	50.6	1805	1 HMW2_MYCGE	P47460 mycoplasma
15	39	50.6	1853	1 MYSA_MOUSE	Q99104 mus musculu
16	38	49.4	305	1 Y861_SYNY3	P73585 synecocyst
17	38	49.4	449	1 YBB9_SCHPO	O60071 schizosacch
18	38	49.4	503	1 TCMO_PEA	Q43067 pisum sativ
19	38	49.4	611	1 VIE3_MCMVS	P29832 murine cyto
20	38	49.4	715	1 CLPB_MYCPN	P75247 mycoplasma
21	38	49.4	1982	1 CHDM_DROME	O97159 drosophila
22	37	48.1	103	1 CYTA_RAT	P01039 rattus norv
23	37	48.1	121	1 YCX6_PORPU	P51216 porphyra pu
24	37	48.1	194	1 VOY8_CAEEL	P34655 caenorhabdi
25	37	48.1	293	1 Y666_METTH	O26762 methanobact
26	37	48.1	408	1 LA_HUMAN	P05455 homo sapien
27	37	48.1	421	1 TYPH_MYCGE	P47297 mycoplasma
28	37	48.1	428	1 U2R1_MOUSE	Q64707 mus musculu
29	37	48.1	457	1 MESE_LEUMA	Q10419 leuconostoc
30	37	48.1	555	1 NFL_CYPJA	Q02916 coturnix co
31	37	48.1	566	1 Y397_MYCGE	P47637 mycoplasma
32	37	48.1	599	1 RF2P_DROME	P14199 drosophila
33	37	48.1	785	1 MUS2_BACSU	P94545 bacillus su

#### RESULT 1

ID	MYS2_DICDI	STANDARD;	PRT;	2116 AA.
AC	P08799;			
DT	01-NOV-1988	(Rel. 09, Created)		
DT	01-OCT-1989	(Rel. 12, Last sequence update)		
DT	15-JUL-1998	(Rel. 36, Last annotation update)		
DE	MYOSIN II HEAVY CHAIN, NON MUSCLE.			
GN	MHCA.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.			
OX	NCBI_TaxID=44689;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87092266; PubMed=3540939;			
RA	Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;			
RT	"Conserved protein domain in a myosin heavy chain gene from Dictyostelium discoideum.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).			
RN	[2]			
RP	PHOSPHORYLATION SITES, AND MUTAGENESIS.			
RC	STRAIN=AX2;			
RX	MEDLINE=90353583; PubMed=2387408;			
RA	Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,			
RL	Gerisch G.;			
RT	"Replacement of threonine residues by serine and alanine in a phosphorylatable heavy chain fragment of Dictyostelium myosin II.";			
RL	FEBS Lett. 269:239-243(1990).			
RN	[3]			
RP	PHOSPHORYLATION SITES.			
RX	MEDLINE=88112226; PubMed=2828113;			
RA	Wagie G., Noegel A., Scheel J., Gerisch G.;			
RT	"Phosphorylation of threonine residues on cloned fragments of the Dictyostelium myosin heavy chain.";			
RL	FEBS Lett. 227:71-75(1988).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.			
RX	MEDLINE=95345067; PubMed=7619795;			
RA	Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,			
RT	Rayment I.;			
RT	"X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.BeFx and MgADP.ALFX.";			
RL	Biochemistry 34:8960-8972(1995).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.			
RX	MEDLINE=95345067; PubMed=7619796;			
RA	Smith C.A., Rayment I.;			
RT	"X-ray structure of the magnesium(II)-pyrophosphate complex of the truncated head of Dictyostelium discoideum myosin to 2.7-A resolution.";			
RL	Biochemistry 34:8973-8981(1995).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.			
RX	MEDLINE=96206189; PubMed=8611530;			
RA	Smith C.A., Rayment I.;			
RT	"X-ray structure of the magnesium(II).ADP.vanadate complex of the			

O9U3V9 drosophila  
P45387 haemophilus  
P23386 saccharomyc  
P12270 homo sapien  
P43720 haemophilus  
P09940 pichia jadi  
Q9PQI7 ureaplasma  
P09413 anabaena fl  
O66612 aquifex aeo  
P08672 plasmodium  
P08675 plasmodium  
P08673 plasmodium

Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";  
Biochemistry 35:5404-5417(1996).  
[7]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.  
RX MEDLINE-97452580; PubMed-9305951;  
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;  
"X-ray structures of the MgADP, MgATPgammA, and MgAMPPNP complexes  
of the Dictyostelium discoideum myosin motor domain.";  
Biochemistry 36:11619-11628(1997).  
[8]  
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.  
RX MEDLINE-98070605; PubMed-9405148;  
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;  
"X-ray crystal structure and solution fluorescence characterization  
of Mg<sub>2</sub>(3')-O-(N-methylanthraniloyl) nucleotides bound to the  
Dictyostelium discoideum myosin motor domain.";  
J. Mol. Biol. 274:394-407(1997).  
RL J. FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE  
ACTIVITY THAT IS ACTIVATED BY ACTIN.  
CC -| SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES  
CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI  
CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS  
CC (MLC-2).  
CC -| SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL  
CC CORTEX.  
CC -| DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
CC MEROMYOSIN (1MM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER  
CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
CC SUBFRAGMENT (S2).  
CC -| DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -| PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES  
CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.  
CC -| MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE  
CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1  
CC POSITION (698).  
CC -| SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF  
CC SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE  
CC HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT  
CC THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE  
CC THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.  
CC -| SIMILARITY: CONTAINS 1 IQ DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M14628; AAA33227.1; -  
PIR; A26655; A26655.  
DR PIR; S00250; S00250.  
DR PDB; 1MMA; 03-DEC-97.  
DR PDB; 1MMD; 17-AUG-96.  
DR PDB; 1MMG; 03-DEC-97.  
DR PDB; 1MMN; 03-DEC-97.  
DR PDB; 1MND; 17-AUG-96.  
DR PDB; 1MNE; 17-AUG-96.  
DR PDB; 1VOM; 23-DEC-96.  
DR PDB; 1LVK; 28-JAN-98.  
DR DictyDb; DD01008; mhca.  
DR InterPro; IPR000048; -  
DR InterPro; IPR001609; -  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF00063; myosin\_head; 1.  
DR PRINTS; PY00193; MYOSINHEAVY.  
KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;  
KW Heptad repeat pattern; Methylation; Alkylation; Phosphorylation.  
FT DOMAIN 1 816 GLOBULAR HEAD (S1).  
DOMAIN 817 2116 RODLIKE TAIL (S2 AND LMM DOMAINS).

FT DOMAIN 817 2116 COILED COIL (POTENTIAL).  
FT NP\_BIND 179 186 ATP.  
FT DOMAIN 638 660 ACTIN-BINDING.  
FT DOMAIN 738 752 ACTIN-BINDING.  
FT MOD\_RES 130 130 METHYLATION (DI-) (POTENTIAL).  
FT MOD\_RES 678 678 ALKYLATION (SH-1).  
FT MOD\_RES 1823 1823 PHOSPHORYLATION (BY MHCK).  
FT MOD\_RES 1833 1833 PHOSPHORYLATION (BY MHCK).  
FT MOD\_RES 2029 2029 PHOSPHORYLATION (BY MHCK).  
SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770BBIE56A1 CRC64;

Query Match 58.4%; Score 45; DB 1; Length 2116;  
Best Local Similarity 69.2%; Pred. No. 35;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EAANKQKQELDEI 14  
|||::|  
Db 1339 EAKNKESELDEI 1351

## RESULT 2

UBAL\_SCHPO STANDARD; PRT; 1012 AA.  
ID UBAL\_SCHPO Q9USY9;  
AC O94609; Q9USY9;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DE 01-OCT-2000 (Rel. 40, Last annotation update)  
DE UBIQUITIN-ACTIVATING ENZYME E1 1 (POLY(A)+ RNA TRANSPORT PROTEIN 3).  
GN PR3 OR SPBC1604.21C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
[1]  
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RP MEDLINE-97311754; PubMed-9168469;  
RA Azad A.K., Tani T., Tsuneyoshi S., Urushiyama S.,  
RA Ohshima Y.;  
RT "Isolation and molecular characterization of mRNA transport mutants in  
Schizosaccharomycetes pombe.";  
RL Mol. Biol. Cell 8:825-841(1997).  
[2]  
RN SEQUENCE OF 611-1012 FROM N.A.  
RP STRAIN=972;  
RA Beck A., Reinhardt R., Lyne M., Rajandream M.A., Barrell B.G.;  
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
CC -| FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLATING WITH ATP ITS  
CC CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS  
CC RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN E1, YIELDING  
CC AN UBIQUITIN-E1 THIOL-ESTER AND FREE AMP.  
CC -| PATHWAY: FIRST STEP IN UBIQUITIN CONJUGATION.  
CC -| SUBUNIT: MONOMER (BY SIMILARITY).  
CC -| SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.  
CC -| SIMILARITY: TO E1 IN OTHER SPECIES.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D87259; BAA75198.1; -  
DR EMBL; AL034433; CAA22354.1; -  
DR InterPro; IPR000011; -  
DR InterPro; IPR000127; -  
DR Pfam; PF00899; Thif family; 2.  
DR Pfam; PF02134; UBACT-repeat; 1.  
DR PROSITE; PS00536; UBIQUITIN\_ACTIVAT\_1; FALSE\_NEG.



DR PROSITE: PS00865; UBIQUITIN ACTIVAT 2; 1.  
KW Ubiquitin conjugation: Ligase; Nuclear protein.  
FT ACT SITE 593 593 BY SIMILARITY.  
SQ SEQUENCE 1012 AA; 112948 MW; AB5207808ACC62D CRC64;

Query Match 55.8%; Score 43; DB 1; Length 1012;  
Best Local Similarity 60.0%; Pred. No. 33;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIS 15  
DB 777 ETAANKQKQELKSA 791

RESULT 3  
ID UCR7\_YEAST STANDARD; PRT; 127 AA.  
AC P00128;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 14 KDA PROTEIN (EC 1.10.2.2)  
DE (COMPLEX III SUBUNIT VII).  
GN QCR7 OR CRO1 OR UCR7 OR YDR529C OR D9719.32.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 28583 / FL100;  
RX MEDLINE=84108379; PubMed=6319130;  
RA de Haan M., van Loon A.P.G.M., Kreike J., Vaessen R.T.M.J.,  
RA Grivell L.A.;  
RT "The biosynthesis of the ubiquinol-cytochrome c reductase complex in yeast. DNA sequence analysis of the nuclear gene coding for the 14-kDa subunit.";  
RL Eur. J. Biochem. 138:169-177(1984).  
RN [2]

RP SEQUENCE FROM N.A.  
RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E.,  
RA Berno A., Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,  
RA Hunkle-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,  
RA Mosedale D., Nakabara K., Namath A., Oefner P., Oh C., Petel F.X.,  
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,  
RA Winant A., Yelton M., Botstein D., Davis R.W.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C  
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH  
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS COMPONENT  
CC IS INVOLVED IN REDOX-LINKED PROTON PUMPING.  
CC -!- CATALYTIC ACTIVITY: OX(2) + 2 FERRICYTOCHROME C = Q +  
CC 2 FERROCYTOCHROME C.  
CC -!- SUBUNIT: FUNGI BCI COMPLEX CONTAINS 10 SUBUNITS; 3 RESPIRATORY  
CC SUBUNITS, 2 CORE PROTEINS AND 5 LOW-MOLECULAR WEIGHT PROTEINS.  
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.  
CC -!- SIMILARITY: BELONGS TO THE UQCRCB/QCR7 FAMILY.  
CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE THE UBIQUINONE-BINDING  
CC PROTEIN (QP-C).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X00256; CAA25064.1; -;  
DR EMBL; U33057; AAB64968.1; -;  
DR PIR; A00121; RDBYUN.  
DR SGD; S0002937; QCR7.

KW Mitochondrion; Inner membrane; Electron transport; Respiratory chain;  
KW Oxidoreductase.  
FT CONFLICT 92 92 E -> Q (IN REF. 1).  
SQ SEQUENCE 127 AA; 14565 MW; 1F1BA3DB6C4067B4 CRC64;

Query Match 54.5%; Score 42; DB 1; Length 127;  
Best Local Similarity 69.2%; Pred. No. 5.5;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EAANKQKQELDEI 14  
DB 111 EAAAKEKDELDNI 123

RESULT 4  
ID IF3\_PROVU STANDARD; PRT; 179 AA.  
AC P33319;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE TRANSLATION INITIATION FACTOR IF-3.  
DE INFC.  
GN Proteus vulgaris.  
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Proteus.  
OX NCBI\_TaxID=585;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13315;  
RX MEDLINE=94010248; PubMed=8405963;  
RA Liveris D., Schwartz J.J., Geertman R., Schwartz I.;  
RT "Molecular cloning and sequencing of infC, the gene encoding translation initiation factor IF3, from four enterobacterial species.";  
RL FEMS Microbiol. Lett. 112:211-216(1993).  
CC -!- FUNCTION: IF-3 BINDS TO THE 30S RIBOSOMAL SUBUNIT AND SHIFTS THE  
CC EQUILIBRIUM BETWEEN 70S RIBOSOMES AND THEIR 50S AND 30S SUBUNITS IN  
CC FAVOR OF THE FREE SUBUNITS, THUS ENHANCING THE AVAILABILITY OF 30S  
CC SUBUNITS ON WHICH PROTEIN SYNTHESIS INITIATION BEGINS.  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- SIMILARITY: BELONGS TO THE IF-3 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L11257; AAC36812.1; -;  
DR HSSP; P02999; 2IFE.  
DR InterPro; IPR001288; -;  
DR Pfam; PF00707; IF3; 1.  
DR PROSITE; PS00938; IF3; 1.  
KW Initiation factor; Protein biosynthesis.  
FT SITE 107 107  
FT IMPORTANT FOR 30S BINDING  
FT (BY SIMILARITY).  
FT SITE 110 110  
FT IMPORTANT FOR 30S BINDING  
FT (BY SIMILARITY).  
SQ SEQUENCE 179 AA; 20535 MW; CA0650013E76D71D CRC64;

Query Match 53.2%; Score 41; DB 1; Length 179;  
Best Local Similarity 46.7%; Pred. No. 11;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 EAANKQKQELDEIST 16  
DB 143 EVLNRIQDLDELAT 157

## RESULT 5

DNAX\_RICPR STANDARD; PRT; 627 AA.  
ID Q9ZDX9;  
AC 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE DNAX PROTEIN (HEAT SHOCK PROTEIN 70) (HSP70).  
GN DNAX OR RP185.  
OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsia; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MADRID E;  
RX MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria.";  
RL Nature 396:133-140(1998).  
CC -!- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
CC -!- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AJ235270; CAAL14651.1;  
DR InterPro; IPR001023;  
DR Pfam; PF00012; HSP70; 1.  
DR PRINTS; PR00301; HEATSHOCK70.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
KW Chaperone; ATP-binding; Heat shock.  
SQ SEQUENCE 627 AA; 68383 MW; A638A0A5A5C8ACC7 CRC64;

Query Match 53.2%; Score 41; DB 1; Length 627;  
Best Local Similarity 56.2%; Pred. No. 42;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 OBAANKQKQELDEIST 16

|||||:|:|:|

Db 257 KEAEKAKKELSSST 272

## RESULT 6

Y328\_MYCPN STANDARD; PRT; 1033 AA.  
ID Y328\_MYCPN  
AC P75310;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL PROTEIN M328 HOMOLOG (P01\_ORF1033).  
GN MPN474 OR MP367.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ATCC 29342 / M129;

RX MEDLINE=97105885; PubMed=8948633;  
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
RA Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
RT pneumoniae.";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF000035; AAB96015.1;  
KW Hypothetical protein.  
SQ SEQUENCE 1033 AA; 118078 MW; B32A330BEA4869BA CRC64;

Query Match 53.2%; Score 41; DB 1; Length 1033;  
Best Local Similarity 64.3%; Pred. No. 70;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps

Qy 1 QEAANKQKQELDEI 14

|||:|:|:|

Db 832 QESYNKVKAELENEI 845

## RESULT 7

M6\_STRPY STANDARD; PRT; 483 AA.  
ID M6\_STRPY  
AC P08089;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE M PROTEIN, SEROTYPE 6 PRECURSOR.  
GN EMW6.  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86111835; PubMed=3511046;  
RA Hollingshead S.K., Fischetti V.F., Scott J.R.;  
RT "Complete nucleotide sequence of type 6 M protein of the group A  
RT Streptococcus. Repetitive structure and membrane anchor.";  
RL J. Biol. Chem. 261:1677-1686(1986).  
RN [2]  
RP SEQUENCE OF 43-122 FROM N.A.  
RX MEDLINE=85166224; PubMed=3885219;  
RA Scott J.R., Pulliam W.M., Hollingshead S.K., Fischetti V.A.;  
RT "Relationship of M protein genes in group A streptococci.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:1822-1826(1985).  
RN [3]  
RP REVIEW.  
RX MEDLINE=91126460; PubMed=1846974;  
RA Fischetti V.A.;  
RT "Neural vector. Herpes may open the way to gene therapy in neurons.";  
RL Sci. Am. 264:32-39(1991).  
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES  
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF  
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO  
CC PHAGOCYTOSIS.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.  
CC -!- SIMILARITY: TO OTHER M PROTEINS.  
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS  
CC IN THE REGION OF THE MEMBRANE ANCHOR.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M11338; AAA26920.1; -  
PIR; A26297; A26297.  
InterPro; IPR001899; -  
Pfam; PF00746; Gram\_pos\_anchor; 1.  
PRINTS; PR00015; GPOSANCHOR.  
PROSITE; PS00343; GRAM\_POS\_ANCHORING; 1.  
Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;  
KW Transmembrane; Coiled coil; Signal.  
FT SIGNAL 1 42  
FT CHAIN 43 483 M PROTEIN, SEROTYPE 6.  
FT DOMAIN 43 457 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 458 477  
FT DOMAIN 478 483 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 478 483 10 X 7 AA TANDEM REPEATS.  
FT DOMAIN 157 269 4.5 X 25 AA TANDEM REPEATS.  
FT DOMAIN 279 347 TWO DIRECTLY REPEATED 27 AMINO ACID  
BLOCKS SEPARATED BY 15 AMINO ACIDS.  
FT DOMAIN 348 411 HYDROPHILIC.  
FT DOMAIN 412 448 GLY/PRO-RICH (CELL WALL-SPANNING).  
FT DOMAIN 449 454 CONSERVED IN GRAM-POSITIVE COCCI SURFACE  
PROTEINS.  
FT SEQUENCE 483 AA; 53472 MW; 68F87F28DB53A448 CRC64;

Query Match 51.9%; Score 40; DB 1; Length 483;  
Best Local Similarity 43.8%; Pred. No. 45;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 16  
Db 149 EEAANKERENKEAIGT 164

RESULT 8  
MYO\_HUMAN STANDARD; PRT; 1976 AA.  
AC P35860; 01-JUN-1994 (Rel. 29, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MYOSIN HEAVY CHAIN, NONMUSCLE TYPE B (CELLULAR MYOSIN HEAVY CHAIN, TYPE B) (NMHC-B).  
GN MYH10.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96025307; PubMed=7499478;  
RA Phillips C.L., Yamakawa K., Adelstein R.S.;  
RT "Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and analysis of human tissues with isoform-specific antibodies.";  
RL J. Muscle Res. Cell Motil. 16:379-389(1995).  
RN [2]  
RP SEQUENCE OF 63-722 FROM N.A.  
RX MEDLINE=91316803; PubMed=1860190;  
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K., Gdula D., Adelstein R.S., Weir L.;  
RT "Human nonmuscle myosin heavy chains are encoded by two genes located on different chromosomes.";  
RL Circ. Res. 69:530-539(1991).  
CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS, CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND CAPING.  
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC

-1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
-1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).  
-1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST TO OTHER NONMUSCLE MYOSINS.  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M69181; AAA99177.1; -  
PIR; B61231; B61231.  
HSP; P08799; 1MND.  
MIM; 160776; -  
InterPro; IPR000048; -  
InterPro; IPR001609; -  
InterPro; IPR002928; -  
Pfam; PF00612; IQ; 1.  
Pfam; PF01576; Myosin\_tail; 1.  
Pfam; PF00063; myosin\_head; 1.  
PRINTS; PR00193; MYOSINHEAVY.  
KW Myosin; Coiled coil; Actin-binding; Alkylation; ATP-binding;  
KW Multigene family.  
FT DOMAIN 1 844 GLOBULAR HEAD (S1).  
FT DOMAIN 845 1976 RODLIKE TAIL (S2 AND LMW DOMAINS).  
FT DOMAIN 845 1976 COILED COIL (POTENTIAL).  
FT NP\_BIND 178 185 ATP (POTENTIAL).  
FT MOD\_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).  
FT MOD\_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).  
SQ SEQUENCE 1976 AA; 228938 MW; B2BB87FF35EA124F CRC64;

Query Match 51.9%; Score 40; DB 1; Length 1976;  
Best Local Similarity 53.3%; Pred. No. 2e+02;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAANKQKQELDEIST 16  
Db 1280 EKASKLQNELDNVST 1294

RESULT 9  
CPI2\_PIG STANDARD; PRT; 35 AA.  
ID CPI2\_PIG  
AC P80736; 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE LEUCOCYTE CYSTEINE PROTEINASE INHIBITOR 2 (PLCPII) (STEFIN D2) (FRAGMENT).  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Thymus;  
RX MEDLINE=97053662; PubMed=8898076;  
RA Lenarcic B., Krizaj I., Zuncic P., Turk V.;  
RT "Differences in specificity for the interactions of stefins A, B and D with cysteine proteinases.";  
RL FEBS Lett. 395:113-118(1996).  
CC -1- FUNCTION: POTENT INHIBITOR OF CATHEPSINS L AND S, AND PAPAIN.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE TYPE 1 CYSTATIN FAMILY.

```

DR HSP; P01040; ICDU.
DR Pfam; PF00031; cystatin; 1.
KW Thiol protease inhibitor.
FT NON_TER 1
FT ACT_SITE 3 3 REACTIVE SITE (BY SIMILARITY).
FT NON_TER 35 35
SQ SEQUENCE 35 AA; 3925 MW; 6A3A4167907D9B94 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 35;
Best Local Similarity 61.5%; Pred. No. 4.2;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QEAANKKQELDE 13
DB 16 QEIAKVKPQLEE 28

RESULT 10
CPIL_PIG STANDARD; PRT; 103 AA.
AC P35479;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE LEUKOCYTE CYSTEINE PROTEINASE INHIBITOR 1 (PLCPI) (STEFIN D1).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP TISSUE=Leukocyte;
RC MEDLINE=94085624; PubMed=8262248;
RA Lenarcic B., Ritonja A., Dolenc I., Stoka V., Berbic S., Pungercar J.,
RA Strukelj B., Turk V.;
RT "Pig leukocyte cysteine proteinase inhibitor (PLCPI), a new member of
RT the stefin family.";
RL FEBS Lett. 336:289-292(1993).
RN [2]
RP SEQUENCE.
RC TISSUE=Thymus;
RX MEDLINE=97053662; PubMed=8898076;
RA Lenarcic B., Krizaj I., Zuncic P., Turk V.;
RT "Differences in specificity for the interactions of stefins A, B and
RT D with cysteine proteinases.";
RL FEBS Lett. 395:113-118(1996).
CC -!- FUNCTION: POTENT INHIBITOR OF CATHEPSINS L AND S, AND PAPAINE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE TYPE 1 CYSTATIN FAMILY.
DR PIR; S40455; S40455.
DR HSP; P01040; ICDU.
DR InterPro; IPR000010; -.
DR Pfam; PF00031; cystatin; 1.
DR PRINTS; PR00295; STEFINA.
DR PROSITE; PS00287; CYSTATIN; 1.
KW Thiol protease inhibitor.
FT MOD_RES 1 9 BLOCKED (PARTIAL).
FT ACT_SITE 9 9 REACTIVE SITE (BY SIMILARITY).
FT SITE 51 55 SECONDARY AREA OF CONTACT.
SQ SEQUENCE 103 AA; 11767 MW; A28B0533D274A55D CRC64;

Query Match 50.6%; Score 39; DB 1; Length 103;
Best Local Similarity 61.5%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QEAANKKQELDE 13
DB 22 QEIAKVKPQLEE 34

```

```

RESULT 11
CYT5_PIG STANDARD; PRT; 103 AA.
AC Q28986;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE CYSTATIN A5 (STEFIN A5).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Pungercar J., Strukelj B.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IS AN INTRACELLULAR THIOL PROTEINASE INHIBITOR.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE TYPE 1 CYSTATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U41732; AAB03261.1; -.
DR HSP; P01040; ICDU.
DR InterPro; IPR000010; -.
DR Pfam; PF00031; cystatin; 1.
DR PRINTS; PR00295; STEFINA.
DR PROSITE; PS00287; CYSTATIN; 1.
KW Thiol protease inhibitor.
FT ACT_SITE 10 10 REACTIVE SITE (BY SIMILARITY).
FT SITE 51 55 SECONDARY AREA OF CONTACT.
SQ SEQUENCE 103 AA; 11743 MW; 9F4FDE8580AAEB44 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 103;
Best Local Similarity 61.5%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QEAANKKQELDE 13
DB 22 QEIAKVKPQLEE 34

RESULT 12
CYT8_PIG STANDARD; PRT; 103 AA.
AC Q28987;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE CYSTATIN A8 (STEFIN A8).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA Pungercar J., Strukelj B.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IS AN INTRACELLULAR THIOL PROTEINASE INHIBITOR.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE TYPE 1 CYSTATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U41733; AAB03262.1; -  
DR HSSP; P01040; 1CYU.  
DR InterPro; IPR000010; -  
DR Pfam; PF00031; Cystatin; 1.  
DR PRINTS; PR00295; STERINA.  
DR PROSITE; PS00287; CYSTATIN; 1.  
KW Thiol protease inhibitor.  
FT ACT\_SITE 10 10 REACTIVE SITE (BY SIMILARITY).  
FT SITE 51 55 SECONDARY AREA OF CONTACT.  
SQ SEQUENCE 103 AA; 11713 MW; 9F521E8580AAB44 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 103;  
Best Local Similarity 61.5%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDE 13  
|| ||| | : : :  
DB 22 QETANKVKQLEE 34

RESULT 13  
ID GAG\_SFV3L STANDARD; PRT; 643 AA.  
AC P27400;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE GAG POLYPROTEIN (CORE POLYPROTEIN).  
GN GAG.  
OS Simian foamy virus (type 3 / strain LK3) (SFV-3).  
OC Viruses; Retroid viruses; Retroviridae; Spumavirus.  
OX NCBI\_TaxID=11644;  
RN [1]

CC SEQUENCE FROM N.A.  
CC MEDLINE-92124734; PubMed-1310187;  
RA Renne R., Friedl E., Schweizer M., Fleps U., Turek R.,  
RA Neumann-Haefelin D.;  
RT "Genomic organization and expression of simian foamy virus type 3  
RT (SFV-3).";  
RL Virology 186:597-608(1992).  
CC -I- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; W74895; AAA47795.1; ALT\_INIT.  
DR PIR; A40820; FOLJLK.  
KW Core protein; Polyprotein.  
SQ SEQUENCE 643 AA; 69785 MW; C53A0575BA9B5949 CRC64;

Query Match 50.6%; Score 39; DB 1; Length 643;  
Best Local Similarity 57.1%; Pred. No. 87;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEI 14  
|| | | : : : :  
DB 127 QETAKQKQKQELDEI 140

RESULT 14  
ID HMW2\_MYCSE STANDARD; PRT; 1805 AA.  
AC P47460;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTADHERENCE HIGH MOLECULAR WEIGHT PROTEIN 2 (CYTADHERENCE ACCESSORY  
DE PROTEIN 2).  
DE HMW2 OR MG218.  
GN HMW2 MYCOPLASMA GENITALIUM.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]

CC SEQUENCE FROM N.A.  
CC STRAIN-ATCC 33530 / G-37;  
CC MEDLINE-96026346; PubMed-7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Eriichman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403(1995).  
RN [2]

CC SEQUENCE OF 557-659 FROM N.A.  
CC STRAIN-ATCC 33530 / G-37;  
CC MEDLINE-94075230; PubMed-8253680;  
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random  
RT sequencing.";  
RL J. Bacteriol. 175:7918-7930(1993).  
CC -I- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH  
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS  
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW  
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS  
CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY  
CC SIMILARITY).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U39701; AAC71437.1; -  
DR EMBL; U02165; AAD12447.1; -  
DR TIGR; MG218; -  
KW Cytoadherence; Structural protein; Coiled coil.  
FT DOMAIN 28 838 COILED COIL (POTENTIAL).  
FT DOMAIN 914 1591 COILED COIL (POTENTIAL).  
FT DOMAIN 1632 1723 COILED COIL (POTENTIAL).  
FT DOMAIN 1777 1804 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1805 AA; 216252 MW; 11D093AF173284FD CRC64;

Query Match 50.6%; Score 39; DB 1; Length 1805;  
Best Local Similarity 46.2%; Pred. No. 2.6e+02;  
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 2 EAANKQKQELDEI 14  
:: |||::|||:  
DB 1512 DSTNKKQKQELNEL 1524

RESULT 15  
ID MY5A\_MOUSE STANDARD; PRT; 1853 AA.

AC Q99104;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE DILUTE MYOSIN HEAVY CHAIN, NON-MUSCLE (MYOSIN VA) (MYOSIN 5A).  
GN MYOSA OR DILUTE.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=91141583; PubMed=1996138;  
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,  
RA Jenkins N.A.;  
RT "Novel myosin heavy chain encoded by murine dilute coat colour  
locus".  
RL Nature 349:709-712(1991).  
RN [2]  
RP REVISIONS.  
RA Mercer J.A., Seperack P.K., Strobel M.C., Copeland N.G.,  
RA Jenkins N.A.;  
RL Nature 352:547-547(1991).  
CC -!- FUNCTION: MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR  
CC ALTERNATIVELY, IT MAY BE REQUIRED FOR SOME POLARIZATION PROCESS  
CC INVOLVED IN DENDRITE FORMATION.  
CC -!- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE  
CC CALMODULIN OR MYOSIN LIGHT CHAINS.  
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. CLASS-5  
CC MYOSIN SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X57377; CAA40651.1; -  
CC PIR; A46761; A46761.  
CC HSP; P08799; IMND.  
CC MGD; MGI:I05976; Myo5a.  
CC InterPro; IPR000048; -  
CC InterPro; IPR001609; -  
CC InterPro; IPR002710; -  
CC Pfam; PF01843; DIL; 1.  
CC Pfam; PF00612; IQ; 6.  
CC Pfam; PF00063; myosin\_head; 1.  
CC PRINTS; PR00193; MYOSINHEAVY.  
CC PROSITE; PS50096; IQ; 5.  
CC Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;  
KW Coiled coil; Phosphorylation.  
FT DOMAIN 1 765 HEAD.  
FT DOMAIN 766 913 NECK.  
FT DOMAIN 914 1040 COILED COIL.  
FT DOMAIN 1041 1853 TAIL.  
FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).  
FT DOMAIN 766 887 5 X APPROXIMATE TANDEM REPEATS.  
FT REPEAT 766 790 1.  
FT REPEAT 791 815 2.  
FT REPEAT 816 837 3.  
FT REPEAT 838 863 4.  
FT REPEAT 864 888 5.  
FT DOMAIN 773 787 CALMODULIN-BINDING (BY SIMILARITY).  
FT DOMAIN 796 810 CALMODULIN-BINDING (BY SIMILARITY).  
FT DOMAIN 821 835 CALMODULIN-BINDING (BY SIMILARITY).  
FT DOMAIN 844 862 CALMODULIN-BINDING (BY SIMILARITY).  
FT DOMAIN 869 884 CALMODULIN-BINDING (BY SIMILARITY).  
FT DOMAIN 892 906 CALMODULIN-BINDING (BY SIMILARITY).  
FT NP\_BIND 163 170 ATP (POTENTIAL).  
FT MOD\_RES 1758 1758 PHOSPHORYLATION (POTENTIAL).

SO SEQUENCE 1853 AA; 215594 MW; 503E93D48CA6B766 CRC64;  
Query Match 50.6%; Score 39; DB 1; Length 1853;  
Best Local Similarity 50.0%; Pred.No. 2.6e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 QEAANKOKQELDEI 14  
:|:|:|:|:|:|:  
Db 1010 EERADKYKQETDQL 1023

Search completed: July 5, 2001, 11:51:45  
Job time: 504 sec

Thu Jul 5 13:58:45 2001

us-09-462-480-11.rsp

Page 9

7

THIS PAGE BLANK (USPTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:50:51 ; Search time 123.78 Seconds  
(without alignments)  
17.102 Million cell updates/sec

Title: US-09-462-480-11

Perfect score: 77

Sequence: 1 QEAANKQKQELDEIST 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_16:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_prodent:\*\*
- 12: sp\_unclassified:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_virus:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	100.0	100	2	069739
2	45	58.4	727	5	Q22505
3	45	58.4	1581	4	Q92614
4	45	58.4	1976	6	Q27991
5	45	58.4	1976	11	Q9JLT0
6	43	55.8	681	5	Q9VDE0
7	43	55.8	1339	5	Q9VM06
8	42	54.5	275	14	Q9DIE2
9	42	54.5	1273	14	Q9QSK2
10	42	54.5	2035	11	Q9TMH9
11	41	53.2	176	8	Q9H1L5
12	41	53.2	211	10	Q9M279
13	41	53.2	545	3	Q05842
14	41	53.2	610	3	Q12136
15	41	53.2	746	5	Q25561
16	41	53.2	829	10	Q9LPI3
17	41	53.2	1410	4	Q14221
18	41	53.2	1411	4	Q15075
19	41	53.2	2007	13	Q02015

09sc28 arabidopsis  
Q9n8a0 plasmodium  
Q164525 arabidopsis  
Q9xja8 streptococ  
Q9su52 arabidopsis  
Q64282 streptococ  
Q64293 streptococ  
Q92576 homo sapien  
Q9nq16 homo sapien  
Q9u145 homo sapien  
Q9w010 drosophila  
Q9zrh5 lactobacill  
Q54587 streptococ  
Q12481 saccharomyc  
Q9euf2 escherichia  
Q91xz7 arabidopsis  
Q9kab2 bacillus ba  
Q9wzb3 drosophila  
O08986 cricetus  
Q45038 schistosoma  
Q19727 caenorhabdi  
O12003 chimpanzee  
Q92603 homo sapien  
Q9nq10 homo sapien  
Q9nq11 homo sapien  
Q9wu41 mus musculu

20 51.9 181 10 09SC28  
21 40 51.9 452 5 09N8A0  
22 40 51.9 699 10 064525  
23 40 51.9 1278 9 09XJA8  
24 40 51.9 1528 10 09SU52  
25 40 51.9 1560 9 064282  
26 40 51.9 1626 9 064293  
27 40 51.9 1723 4 092576  
28 40 51.9 1957 4 09NQ16  
29 40 51.9 2039 4 09U145  
30 40 51.9 2263 5 09W010  
31 39 50.6 85 2 09ZRH5  
32 39 50.6 150 2 054587  
33 39 50.6 300 3 012481  
34 39 50.6 309 2 09EUF2  
35 39 50.6 434 10 091XZ7  
36 39 50.6 521 2 09KAB2  
37 39 50.6 599 5 09WZB3  
38 39 50.6 628 11 008986  
39 39 50.6 648 5 045038  
40 39 50.6 775 5 019727  
41 39 50.6 874 14 012003  
42 39 50.6 1152 4 092603  
43 39 50.6 1154 4 09NQ10  
44 39 50.6 1185 4 09NQ11  
45 39 50.6 1202 11 09WU41

#### ALIGNMENTS

RESULT 1  
069739 PRELIMINARY; PRT; 100 AA.  
ID O69739  
AC O69739;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DE HYPOTHETICAL 10.8 KDA PROTEIN.  
GN LHP OR RV3874 OR MTV027.09.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RA "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX Berthet F.-X., Birk Rasmussen P., Andersen P., Cicquel B.;  
RA "Promoter analysis of the M. tuberculosis orfIC gene encoding the  
early secreted antigenic target 6 kDa (ESAT-6).";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL021210; CAAL17966.1; -  
DR EMBL; AF004671; AAC83445.1; -  
DR TubercuList; RV3874; -  
KW Hypothetical protein.  
SQ SEQUENCE 100 AA; 10794 MW; 285F4FC96F55D194 CRC64;

Query Match 100.0%; Score 77; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 0.0001;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 16  
DB 59 QEAANKQKQELDEIST 74

RESULT 2  
Q22505 PRELIMINARY; PRT; 727 AA.  
ID Q22505;  
AC Q22505;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE SIMILAR TO L96 PROTEIN.  
GN TI4G11.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
RA Smailson N., Smith A., Sonhammer E., Staden K., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans".  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Fulton B., Hawkins J.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Waterston R.;  
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U41024; AAA82345.1; -;  
SQ SEQUENCE 727 AA; 81325 MW; 42581312925DF1F9 CRC64;

Query Match 58.4%; Score 45; DB 5; Length 727;  
Best Local Similarity 60.0%; Pred. No. 50;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIS 15  
DB 313 QETANKLHQELDEIN 327

RESULT 3  
Q92614 PRELIMINARY; PRT; 1581 AA.  
ID Q92614;  
AC Q92614;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE MYELOBLAST KIAA0216.  
GN KIAA0216.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Calarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=BONE MARROW;  
RX MEDLINE=97191544; PubMed=9039502;  
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,  
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. VI.  
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
RT analysis of cDNA clones from cell line KG-1 and brain.";  
RL DNA Res. 3:321-329(1996).  
DR EMBL; D86970; BAA13206.1; -;  
DR HSP; P08799; IMMN.  
DR InterPro; IPR000048; -;  
DR InterPro; IPR001553; -;  
DR InterPro; IPR001609; -;  
DR Pfam; PF00063; myosin\_head; 1.  
DR Pfam; PF00612; IQ; 1.  
DR ProDom; PD000355; -; 1.  
DR SMART; SM00015; IQ; 1.  
SQ SEQUENCE 1581 AA; 180549 MW; FFA85576ABFC56B3 CRC64;

Query Match 58.4%; Score 45; DB 4; Length 1581;  
Best Local Similarity 75.0%; Pred. No. 1e+02;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps

QY 2 EAANKQKQELDE 13  
DB 1326 EAANKQKQELQE 1337

RESULT 4  
Q27991 PRELIMINARY; PRT; 1976 AA.  
ID Q27991;  
AC Q27991;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-AUG-1999 (TREMBlrel. 11, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE NONMUSCLE MYOSIN HEAVY CHAIN B.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ohara M., Ishiguro N., Shinagawa M.;  
RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 204-302 FROM N.A.  
RC TISSUE=CEREBRAL CORTEX;  
RX MEDLINE=95301542; PubMed=7782316;  
RA Itoh K., Adelstein R.S.;  
RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle  
RT myosin heavy chain II-B.";  
RL J. Biol. Chem. 270:14533-14540(1995).  
DR EMBL; AB022023; BAA36494.1; -;  
DR EMBL; U15716; AAA87715.1; -;  
DR HSP; P08799; ILVK.  
DR InterPro; IPR000048; -;  
DR InterPro; IPR001609; -;  
DR InterPro; IPR002017; -;  
DR InterPro; IPR002928; -;  
DR Pfam; PF00063; myosin\_head; 1.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF01576; Myosin\_tail; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR ProDom; PD000355; -; 1.  
SQ SEQUENCE 1976 AA; 229097 MW; 6144354451C0F790 CRC64;

Query Match 58.4%; Score 45; DB 6; Length 1976;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAANKOKQELDEIST 16  
| | | | | : | | | : | |  
Db 1280 EKANKLQNELDNVST 1294

RESULT 5

ID Q9JLT0 PRELIMINARY; PRT; 1976 AA.

AC Q9JLT0; 2000 (TREMELREL. 15, Created)

DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)

DT 01-MAR-2001 (TREMELREL. 16, Last annotation update)

DE NONMUSCLE MYOSIN HEAVY CHAIN-B.

OS Rattus norvegicus (Rat)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;

RA Tam J.W.P., Chan K.W., Hsiao W.L.W.;

RT "Suppression of transformed phenotype by overexpression of nonmuscle  
myosin heavy chain-B in p53val135-transformed rat fibroblast.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF139055; AAF61445.1; -

DR InterPro: IPR000048; -

DR InterPro: IPR001609; -

DR InterPro: IPR002017; -

DR InterPro: IPR002928; -

DR Pfam: PF00063; myosin\_head; 1.

DR Pfam: PF00612; IQ; 1.

DR Pfam: PF01576; Myosin\_tail; 1.

DR PRINTS: PR00193; MYOSINHEAVY.

DR SMART: SM00015; IQ; 1.

SQ SEQUENCE 1976 AA; 228963 MW; E32708BF9BF2B470 CRC64;

Query Match 58.4%; Score 45; DB 11; Length 1976;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EAANKOKQELDEIST 16  
| | | | | : | | | : | |  
Db 1280 EKANKLQNELDNVST 1294

RESULT 6

ID Q9VDE0 PRELIMINARY; PRT; 681 AA.

AC Q9VDE0; 2000 (TREMELREL. 13, Created)

DT 01-MAY-2000 (TREMELREL. 13, Created)

DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)

DT 01-MAR-2001 (TREMELREL. 16, Last annotation update)

DE CG10825 PROTEIN.

GN CG10825.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Anatolides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL: AF003733; AAF55856.1; -

DR FlyBase: FBgn003860; CG10825.

DR InterPro: IPR000980; -

DR SMART: SM00252; SH2; 1.

SQ SEQUENCE 681 AA; 78529 MW; FF140F97D5252699 CRC64;

Query Match 55.8%; Score 43; DB 5; Length 681;  
Best Local Similarity 64.3%; Pred. No. 94;  
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 AANKOKQELDEIST 16  
| | | | | : | | | : | |  
Db 667 AANKKIELEKIMT 680

RESULT 7

ID Q9VM06 PRELIMINARY; PRT; 1339 AA.

AC Q9VM06; 2000 (TREMELREL. 13, Created)

DT 01-MAY-2000 (TREMELREL. 13, Created)

DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)

DT 01-JUN-2000 (TREMELREL. 14, Last annotation update)

DE CG9279 PROTEIN.

GN CG9279.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Anatolides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.A., Bouck J., Brokstein P., Brottier P.,  
 RA Burris K.C., Busam D.R., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J., Yao Q.A.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003516; AAF9149.1;  
 DR FlyBase; FBgn0036882; CG9279.  
 DR InterPro; IPR000938;  
 DR Pfam; PF01302; CAP\_GLY; 1.  
 DR PROSITE; PS00845; CAP\_GLY; 1.  
 SQ SEQUENCE 1339 AA; 152535 MW; 3BB624E7EBC7C0B6 CRC64;

Query Match 55.8%; Score 43; DB 5; Length 1339;  
 Best Local Similarity 53.3%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDEIST 15  
 | : : : : :  
 Db 435 QDVSSKRLKRELEITA 449

RESULT 8  
 Q9DIE2 PRELIMINARY; PRT; 275 AA.  
 AC Q9DIE2;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE DNA POLYMERASE (FRAGMENT).  
 GN POL.  
 OS Iridovirus CIV.  
 OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus.  
 OX NCBI\_TaxID=113375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20540044; PubMed=11086137;  
 RA Stasiak K., Demattè M.V., Federici B.A., Bigot Y.;  
 RT "Phylogenetic position of the *Diadromus pulchellus* ascovirus DNA  
 RT polymerase among viruses with large double-stranded DNA genomes.";  
 RL J. Gen. Virol. 81:3059-3072(2000).  
 DR EMBL; AJ279816; CAC19195.1;  
 FT NON\_TER 275  
 FT SEQUENCE 275 AA; 31746 MW; 0AB60A2FD34131F5 CRC64;

Query Match 54.5%; Score 42; DB 14; Length 275;

Best Local Similarity 46.7%; Pred. No. 58;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 EAANKQKQELDEIST 16  
 | : : : : :  
 Db 191 EKENSKEDIDEITS 205  
 RESULT 9  
 Q9QSK2 PRELIMINARY; PRT; 1273 AA.  
 AC Q9QSK2;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE DNA POLYMERASE.  
 GN DPOL.  
 OS Chilo iridescent virus (CIV) (Insect iridescent virus type 6).  
 OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus.  
 OX NCBI\_TaxID=10488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Muller K., Tidona C.A., Bahr U., Darai G.;  
 RT "The DNA sequence of Chilo iridescent virus between the genome  
 RT coordinates 0.974 and 0.101.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N  
 CC PYROPHOSPHATE + DNA(N) (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.  
 DR EMBL; AF083915; AAD48150.1;  
 DR InterPro; IPR002064;  
 DR Pfam; PF00136; DNA\_pol\_B; 5.  
 DR PRINTS; PR00106; DNAPOLB.  
 DR SMART; SM00486; POLBc; 1.  
 KW DNA replication; DNA-binding; DNA-directed DNA polymerase.  
 SQ SEQUENCE 1273 AA; 147662 MW; CC514AD83F7299A3 CRC64;

Query Match 54.5%; Score 42; DB 14; Length 1273;  
 Best Local Similarity 46.7%; Pred. No. 2.4e+02;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 EAANKQKQELDEIST 16  
 | : : : : :  
 Db 826 EKENSKEDIDEITS 840

RESULT 10  
 Q9JMH9 PRELIMINARY; PRT; 2035 AA.  
 AC Q9JMH9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
 DE MYOSIN CONTAINING PDZ DOMAIN.  
 GN MYSPDZ.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL6;  
 RX MEDLINE=20200096; PubMed=10733906;  
 RA Furusawa T., Ikawa S., Yanai N., Obinata M.;  
 RT "Isolation of a novel PDZ-containing Myosin from hematopoietic  
 RT supportive bone marrow stromal cell lines.";  
 RL Biochem. Biophys. Res. Commun. 270:67-75(2000).  
 DR EMBL; AB026497; BAA93660.1;  
 DR InterPro; IPR000048;  
 DR InterPro; IPR001478;  
 DR InterPro; IPR001609;  
 DR Pfam; PF00063; myosin\_head; 2.

DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00612; IQ; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR SMART; SM00015; IQ; 1. 230905 MW; C5CE4BB053475119 CRC64;  
SQ SEQUENCE 2035 AA; 230905 MW; C5CE4BB053475119 CRC64;

Query Match 54.5%; Score 42; DB 11; Length 2035;  
Best Local Similarity 66.7%; Pred. No. 3.7e+02;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EAANKKQKQELDE 13  
I : I : I : I : I : I :  
Db 1780 EESNKEKQELQ 1791

RESULT 11  
Q9TL15 PRELIMINARY; PRT; 176 AA.  
AC Q9TL15;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE CFO SUBUNIT I OF ATP SYNTHASE.  
GN ATPF.  
OS Nephroselmis olivacea.  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;  
OC Chlorodendraceae; Chlorodendraceae; Nephroselmis.  
OX NCBI\_TaxID=31312;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99398694; PubMed=10468594;  
RA Turmel M., Otis C., Lemieux C.;  
RT "The complete chloroplast DNA sequence of the green alga Nephroselmis  
olivacea: insights into the architecture of ancestral chloroplast  
genomes";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:10248-10253(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Turmel M., Otis C., Lemieux C.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF137379; AAD54801.1; -;  
DR InterPro; IPR001302; -;  
DR InterPro; IPR002146; -;  
DR Pfam; PF00430; ATP-synt\_B; 1.  
DR ProDom; PD003995; -; 1.  
KW Chloroplast.  
SQ SEQUENCE 176 AA; 20113 MW; 0F6731EBF9F11E26 CRC64;

Query Match 53.2%; Score 41; DB 8; Length 176;  
Best Local Similarity 50.0%; Pred. No. 54;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 QEAANKKQKQELDEIST 16  
I : I : I : I : I : I :  
Db 68 QDAVNQKQALQELT 83

RESULT 12  
Q9M279 PRELIMINARY; PRT; 211 AA.  
AC Q9M279;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE HYPOTHETICAL 24.7 KDA PROTEIN.  
GN F21F14.10.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Choisne N., Robert C., Brottier P., Wincker P., Cattolico L.,  
RA Attiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X., Quetier F., Salanoubat M.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL138642; CAB71891.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 211 AA; 24684 MW; 77688F50645CC43F CRC64;

Query Match 53.2%; Score 41; DB 10; Length 211;  
Best Local Similarity 66.7%; Pred. No. 64;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 NKQKQELDEIST 16  
I : I : I : I : I : I :  
Db 28 NKQKQERDMVKT 39

RESULT 13  
Q05842 PRELIMINARY; PRT; 545 AA.  
AC Q05842;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1996 (Tremblrel. 08, Last annotation update)  
DE D1545.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97127830; PubMed=8972581;  
RA Delaveau T., Blugeon C., Jacq C., Perea J.;  
RT "Analysis of a 23 kb region on the left arm of yeast chromosome IV";  
RL Yeast 12:1587-1592(1996).  
DR EMBL; X97751; CAA66339.1; -;  
SQ SEQUENCE 545 AA; 63003 MW; 51455448475962B4 CRC64;

Query Match 53.2%; Score 41; DB 3; Length 545;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EAANKKQKQELDE 13  
I : I : I : I : I : I :  
Db 277 EDGKQKQEIIDE 288

RESULT 14  
Q12136 PRELIMINARY; PRT; 610 AA.  
AC Q12136;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE ORF YDL153C.  
GN SAS10.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Perea J., Blugeon C., Delaveau T., Jacq C.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

RN [2]  
RP SEQUENCE FROM N.A.  
RA MIPS;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kine J.D., Kamakaka R.T.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; 274201; CAA98726.1; -;  
DR EMBL; U63063; AAB05801.1; -;  
DR SGD; S0002312; SAS10  
SQ SEQUENCE 610 AA; 70259 MW; 9278BF873834FC7E CRC64;

Query Match 53.2%; Score 41; DB 3; Length 610;  
Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 EAANKQKQELDE 13  
Db 342 EDGGKQKQEI 353

RESULT 15  
Q25561  
ID Q25561 PRELIMINARY; PRT; 746 AA.  
AC Q25561;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-JUN-2000 (TREMELrel. 14, Last annotation update)  
DE MYOSIN II HEAVY CHAIN (FRAGMENT).  
OS Naegleria fowleri.  
OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.  
OX NCBI\_TaxID=5763;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LEE MP;  
RA Shaw D.R., Sullivan P.K., Marciano-Cabral F., Ennis H.L.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U43192; AAB01786.1; -;  
DR InterPro; IPR002017; -;  
FT NON\_TER 1  
SQ SEQUENCE 746 AA; 87781 MW; 3771A80AAFC2FF43 CRC64;

Query Match 53.2%; Score 41; DB 5; Length 746;  
Best Local Similarity 42.9%; Pred. No. 2.1e+02;  
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 1 QEAANKQKQELDEI 14  
Db 575 EESATQKREIEL 588

Search completed: July 5, 2001, 11:50:52  
Job time: 475 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:45:43 ; Search time 130.35 Seconds  
(without alignments)  
7.441 Million cell updates/sec

Title: US-09-462-480-10

Perfect score: 83

Sequence: 1 GSLOCQWRGAAGTAAA 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601.\*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	83	100.0	16	20	AAV03710	M. tuberculosis LH
2	79	95.2	27	18	AAW32457	Mycobacterium tube
3	79	95.2	27	19	AAW81695	M. tuberculosis im
4	79	95.2	27	20	AAV39125	M. tuberculosis an
5	79	95.2	80	18	AAW32454	Mycobacterium tube
6	79	95.2	80	18	AAW32386	Mycobacterium tube
7	79	95.2	80	19	AAW81707	M. tuberculosis im
8	79	95.2	80	19	AAW64340	Mycobacterium tube
9	79	95.2	80	20	AAV39137	M. tuberculosis an
10	79	95.2	80	20	AAV38994	M. tuberculosis re
11	79	95.2	95	18	AAW32444	Mycobacterium tube

12	79	95.2	95	18	AAW32376	Mycobacterium tube
13	79	95.2	95	19	AAW81747	M. tuberculosis im
14	79	95.2	95	19	AAW64321	Mycobacterium tube
15	79	95.2	95	20	AAV32097	Mycobacterium tube
16	79	95.2	95	20	AAV39118	M. tuberculosis an
17	79	95.2	95	20	AAV38981	M. tuberculosis re
18	79	95.2	100	19	AAW81706	M. tuberculosis im
19	79	95.2	100	19	AAW64339	Mycobacterium tube
20	79	95.2	100	20	AAV39136	M. tuberculosis an
21	79	95.2	100	20	AAV38993	M. tuberculosis re
22	79	95.2	100	20	AAV03705	M. tuberculosis LH
23	79	95.2	100	22	AAW35218	M. tuberculosis RV3
24	79	95.2	100	22	AAW19845	Mycobacterium tube
25	79	95.2	802	19	AAW81746	M. tuberculosis fu
26	79	95.2	802	19	AAW64379	Mycobacterium anti
27	79	95.2	802	20	AAV32063	Mycobacterium tube
28	79	95.2	802	20	AAV39224	M. tuberculosis fu
29	79	95.2	802	20	AAV39176	M. tuberculosis fu
30	79	95.2	802	20	AAV39081	M. tuberculosis fu
31	79	95.2	802	20	AAV39033	M. tuberculosis fu
32	71	85.5	49	20	AAV03706	M. tuberculosis LH
33	52	62.7	16	18	AAW32456	Mycobacterium tube
34	52	62.7	16	19	AAW81694	M. tuberculosis im
35	52	62.7	16	20	AAV39124	M. tuberculosis an
36	48	57.8	103	22	AAW35234	M. tuberculosis RV3
37	46	55.4	593	21	AAW43002	Human ORFX ORF2766
38	42	50.6	31	21	AAV59109	N. clavipes spider
39	42	50.6	34	16	AAW06201	Spider dragline va
40	42	50.6	34	20	AAV33400	N. clavipes spidro
41	42	50.6	34	21	AAV59105	N. clavipes spider
42	42	50.6	34	21	AAV59117	N. clavipes spider
43	42	50.6	101	16	AAW99054	Spider dragline va
44	42	50.6	101	16	AAW99056	Spider dragline va
45	42	50.6	101	16	AAW99052	Spider dragline va

## ALIGNMENTS

RESULT 1

AAV03710

ID AAV03710 standard; Protein; 16 AA.

XX

AC AAV03710;

DT 07-JUN-1999 (first entry)

XX

DE M. tuberculosis LHP polypeptide antigenic fragment.

XX

ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;

KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;

KW immune response.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9904005-A1.

XX

PD 28-JAN-1999.

XX

PF 16-JUL-1998; 98WO-1B01091.

XX

PR 16-JUL-1997; 97US-0052631.

XX

PA (INSP ) INST PASTEUR.

PA (STAT-) STATENS SERUM INST.

XX

PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX

DR WPI; 1999-132249/11.

XX

PT New nucleic acid containing regulator and LHP gene of Mycobacterium

PT tuberculosis - useful in vaccines, for diagnosis, and for expression

PT of heterologous proteins

XX Claim 21; Page 65; 88pp; English.  
 PS The present invention is directed to a polynucleotide carrying the  
 CC regulatory expression signals of the ESAT-6 protein as well as an open  
 CC reading frame coding for an antigenic protein LHP from Mycobacterium  
 CC tuberculosis. Host cells comprising the polynucleotide are used for the  
 CC recombinant expression of the protein. The recombinant polypeptide can  
 CC be used as immunogens and vaccines, to protect against bacteria of the  
 CC M. tuberculosis complex in humans or animals (the vaccines may include  
 CC other immunogenic proteins of the bacteria or their fragments,  
 CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by  
 CC detection of specific antibodies. The regulatory region present in the  
 CC polynucleotide may be used to express almost any heterologous protein in  
 CC mycobacteria, particularly as a fusion with polyhistidine. The two  
 CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to  
 CC provide a synergistic increase in ability to induce a protective immune  
 CC response. Sequences AY03706-713 represent antigenic fragments of the  
 CC LHP polypeptide.  
 XX  
 SQ Sequence 16 AA;

Query Match 100.0%; Score 83; DB 20; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAA 16  
 DB 1 gslgqwrngaagtaaa 16  
 |||||

RESULT 2  
 AAW32457  
 ID AAW32457 standard; Protein; 27 AA.

XX AAW32457;  
 XX 09-JAN-1998 (first entry)  
 DT Mycobacterium tuberculosis antigen Tb38-1 Peptide 3.  
 DE Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M. tuberculosis.  
 XX Mycobacterium tuberculosis.  
 OS  
 XX WO9709428-A2.  
 PN 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.  
 XX 12-JUL-1996; 96US-0680574.  
 PR 01-SEP-1995; 95US-0523436.  
 PR 22-SEP-1995; 95US-0533634.  
 PR 22-MAR-1996; 96US-0620874.  
 PR 05-JUN-1996; 96US-0659683.

XX (CORI-) CORIXA CORP.  
 XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;  
 XX WPI; 1997-192903/17.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
 PT useful in vaccines for prevention or treatment of tuberculosis, also  
 PT for diagnosis  
 XX Example 3; Page 130; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M. tuberculosis  
 CC antigen, Tb38-1 Peptide 3. The immunogenic protein, and fusion proteins  
 CC containing one or more of the proteins or one of the proteins plus  
 CC ESAT-6, are useful in vaccines, preferably when formulated with a  
 CC non-specific adjuvant, to induce an immune response against  
 CC M. tuberculosis (for treatment or prevention).  
 XX  
 SQ Sequence 27 AA;

Query Match 95.2%; Score 79; DB 18; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAA 15  
 DB 7 gslgqwrngaagtaa 21  
 |||||

RESULT 3  
 AAW81695  
 ID AAW81695 standard; Protein; 27 AA.

XX AAW81695;  
 XX 27-JAN-1999 (first entry)  
 DT M. tuberculosis immunogenic polypeptide Tb38-1 peptide 3.  
 DE Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 XX Mycobacterium tuberculosis.  
 OS  
 XX WO9816646-A2.  
 PN 23-APR-1998.  
 PD 07-OCT-1997; 97WO-US18293.  
 PF 13-MAR-1997; 97US-0818112.  
 PR 11-OCT-1996; 96US-0730510.  
 XX (CORI-) CORIXA CORP.  
 XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX WPI; 1998-261042/23.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX Disclosure; Page 122; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.  
 XX  
 SQ Sequence 27 AA;

Query Match 95.2%; Score 79; DB 19; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAA 15



Db 7 gslqgqrgaagtaa 21  
|||||

RESULT 4  
AA39125  
ID AAY39125 standard; peptide; 27 AA.

XX AC AAY39125;

XX DT 05-NOV-1999 (first entry)

XX DE M. tuberculosis antigen Tb38-1 peptide 3 amino acid sequence.

XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
KW immune response; skin test.

XX OS Mycobacterium tuberculosis.

XX PN WO9942076-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03268.

XX PR 05-MAY-1998; 98US-0072967.

XX PR 18-FEB-1998; 98US-0025197.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX DR WPI; 1999-527409/44.

XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions

XX PS Example 3; Page 117; 299pp; English.

XX CC The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to  
CC AAY39225 are used in the exemplification of the present invention.

XX SQ Sequence 27 AA;

Query Match 95.2%; Score 79; DB 20; Length 27;  
Best Local Similarity 100.0%; Pred. No. 6.3e-06; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 0;

QY 1 GSIQGGWRGAAGTAA 15  
Db 7 gslqgqrgaagtaa 21  
|||||

RESULT 5  
AAW32454  
ID AAW32454 standard; Protein; 80 AA.

XX AC AAW32454;

XX DT 09-JAN-1998 (first entry)

XX

DE Mycobacterium tuberculosis antigen Tb38-1F3.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis.

XX OS Mycobacterium tuberculosis.

XX PN WO9709428-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US14674.

XX PR 12-JUL-1996; 96US-0680574.

XX PR 01-SEP-1995; 95US-0523436.

XX PR 22-SEP-1995; 95US-0533634.

XX PR 22-MAR-1996; 96US-0620874.

XX PR 05-JUN-1996; 96US-0659683.

XX PA (CORI-) CORIXA CORP..

XX PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

XX DR WPI; 1997-192903/17.

XX DR N-PSDB; AAT91526.

XX PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
PT useful in vaccines for prevention or treatment of tuberculosis, also  
PT for diagnosis

XX PS Example 3; Page 149; 168pp; English.

XX CC A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb38-1F3. The immunogenic protein, and fusion proteins  
CC containing one or more of the proteins or one of the proteins plus  
CC ESAT-6, are useful in vaccines, preferably when formulated with a  
CC non-specific adjuvant, to induce an immune response against  
CC M.tuberculosis (for treatment or prevention).

XX SQ Sequence 80 AA;

Query Match 95.2%; Score 79; DB 18; Length 80;  
Best Local Similarity 100.0%; Pred. No. 2e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIQGGWRGAAGTAA 15  
Db 17 gslqgqrgaagtaa 31  
|||||

RESULT 6  
AAW32386  
ID AAW32386 standard; Protein; 80 AA.

XX AC AAW32386;

XX DT 13-JAN-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen Tb38-1F3.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis.

XX OS Mycobacterium tuberculosis.

XX PN WO9709429-A2.

XX PD 13-MAR-1997.

```

XX PF 30-AUG-1996; 96WO-US14675.
XX PR 12-JUL-1996; 96US-0680573.
XX PR 01-SEP-1995; 95US-0523435.
XX PR 22-SEP-1995; 95US-0532136.
XX PR 22-MAR-1996; 96US-0620280.
XX PR 05-JUN-1996; 96US-0658800.
XX PA (CORI-) CORIXA CORP.
XX XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
PI Twardzik DR, Vedvick TH;
XX XX WPI; 1997-192904/17.
XX DR N-PSDB; AAT91460.
XX XX
XX PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
XX PT - useful for diagnosis of M. tuberculosis infection
XX XX
XX PS Example 3; Page 163; 190pp; English.
XX CC A new immunogenic polypeptide has been developed comprising an
XX CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
XX CC its variant differing only in conservative substitutions and/or
XX CC modifications). The present sequence represents a M. tuberculosis
XX CC antigen, Tb38-1F3. The immunogenic polypeptide can be used to diagnose
XX CC M. tuberculosis infection by forming complexes with specific
XX CC antibodies in the sample. Fragments of DNA encoding the immunogenic
XX CC polypeptide can be used as diagnostic primers or probes and agents
XX CC that bind to the antigen, especially monoclonal antibodies or
XX CC equivalent polyclonal antibodies, are also used for diagnosis.
XX XX
XX SQ Sequence 80 AA;

Query Match 95.2%; Score 79; DB 18; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAA 15
Db |||||
17 gslggqwrngaagtaa 31

RESULT 7
AAW81707
ID AAW81707 standard; Protein; 80 AA.
XX AC AAW81707;
XX DT 27-JAN-1999 (first entry)
XX XX M. tuberculosis immunogenic polypeptide Tb38-1F3.
XX KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX KW vaccine; pharmaceutical; infection; diagnosis.
XX OS Mycobacterium tuberculosis.
XX PN WO9816646-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18293.
XX XX
XX PR 13-MAR-1997; 97US-0818112.
XX PR 11-OCT-1996; 96US-0730510.
XX PA (CORI-) CORIXA CORP.
XX XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

```

```

XX WPI; 1998-261042/23.
XX DR N-PSDB; AAV64509.
XX XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX PT and for diagnosis, treatment and prevention of tuberculosis
XX PS Example 3B; Page 139-140; 230pp; English.
XX XX
XX CC This sequence represents an immunogenic portion of a soluble
XX CC Mycobacterium tuberculosis (MT) antigen which can be used in a method
XX CC for inducing protective immunity against tuberculosis (TB). This
XX CC sequence can be formulated into vaccines and/or pharmaceutical
XX CC compositions for immunising against M. tuberculosis infection or may
XX CC be used for the diagnosis of tuberculosis.
XX XX
XX SQ Sequence 80 AA;

Query Match 95.2%; Score 79; DB 19; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAA 15
Db |||||
17 gslggqwrngaagtaa 31

RESULT 8
AAW64340
ID AAW64340 standard; Protein; 80 AA.
XX AC AAW64340;
XX DT 09-NOV-1998 (first entry)
XX XX Mycobacterium tuberculosis antigen Tb38-1F3.
XX DE Tuberculosis; infection; diagnosis; antigen; Tb38-1F3.
XX KW Mycobacterium tuberculosis strain H37Rv.
XX OS Mycobacterium tuberculosis strain H37Rv.
XX PN WO9816645-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18214.
XX XX
XX PR 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX PA (CORI-) CORIXA CORP.
XX XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX WPI; 1998-251292/22.
XX DR N-PSDB; AAV44400.
XX XX
XX PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX PT and diagnosis of tuberculosis
XX PS Example 3; Page 146; 250pp; English.
XX XX
XX CC This polypeptide comprises Mycobacterium tuberculosis antigen
XX CC Tb38-1F3. A DNA molecule (see AAV44400) coding for Tb38-1F3 was
XX CC isolated from a M. tuberculosis strain H37Rv genomic library. The
XX CC invention relates to compositions and methods for diagnosing
XX CC tuberculosis. It provides polypeptides (see AAW64291-W64379)
XX CC comprising an antigenic portion of a soluble M. tuberculosis
XX CC antigen, or an immunogenic portion of an M. tuberculosis antigen,

```

CC as well as DNA sequences encoding such polypeptides, recombinant  
 CC expression vectors and transformed or transfected host cells. Also  
 CC claimed are methods and diagnostic kits for detecting M.  
 CC tuberculosis infection in a patient using these polypeptides,  
 CC antibodies or oligonucleotide probes and primers, for the diagnosis  
 CC of tuberculosis.  
 XX  
 XX Sequence 80 AA;

Query Match 95.2%; Score 79; DB 19; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIQGWGRGAAGTAA 15  
 DB 17 GSIQGWGRGAAGTAA 31  
 |||||

RESULT 9  
 AAY39137  
 ID AAY39137 standard; Protein; 80 AA.  
 XX

AC AAY39137;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis antigen Tb38-1F3 amino acid sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test.

XX Mycobacterium tuberculosis.

OS

XX WO9942076-A2.

PN 26-AUG-1999.

PD 17-FEB-1999; 99WO-US03268.

PF 05-MAY-1998; 98US-0072967.

PR 18-FEB-1998; 98US-0025197.

XX (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527409/44.

DR N-PSDB; AA219310.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions

PS Example 3; Page 134-135; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to  
 CC AAY39225 are used in the exemplification of the present invention.

XX Sequence 80 AA;

Query Match 95.2%; Score 79; DB 20; Length 80;

Best Local Similarity 100.0%; Pred. No. 2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSIQGWGRGAAGTAA 15  
 DB 17 GSIQGWGRGAAGTAA 31  
 |||||

RESULT 10  
 AAY38994

ID AAY38994 standard; Protein; 80 AA.

XX

AC AAY38994;

DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen protein Tb38-1F3.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity.

XX Mycobacterium tuberculosis.

XX WO9942118-A2.

PN 26-AUG-1999.

PD 17-FEB-1999; 99WO-US03265.

PF 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527416/44.

DR N-PSDB; AA219098.

XX New polypeptide comprising antigenic portions of M. tuberculosis

PS Example 3; Page 180; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.

XX Sequence 80 AA;

Query Match 95.2%; Score 79; DB 20; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIQGWGRGAAGTAA 15  
 DB 17 GSIQGWGRGAAGTAA 31  
 |||||

RESULT 11  
 AAW32444

ID AAW32444 standard; Protein; 95 AA.

XX

AC AAW32444;

DT 09-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen Tb38-1.

```

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9709428-A2.
XX
XX 13-MAR-1997.
XX
XX
XX 30-AUG-1996; 96WO-US14674.
XX
XX 12-JUL-1996; 96US-0680574.
XX
XX 01-SEP-1995; 95US-0523436.
XX
XX 22-SEP-1995; 95US-0533634.
XX
XX 22-MAR-1996; 96US-0620874.
XX
XX 05-JUN-1996; 96US-0659683.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky VAW;
PI Twardzik DR, Vedvick TH;
XX
XX WPI; 1997-192903/17.
XX
XX N-PSDB; AAT91509.
XX
XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
XX useful in vaccines for prevention or treatment of tuberculosis, also
XX for diagnosis
XX
XX Example 3; Page 124; 168pp; English.
XX
XX A new immunogenic polypeptide has been developed comprising an
XX immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
XX its variant differing only in conservative substitutions and/or
XX modifications). The present sequence represents a M.tuberculosis
XX antigen, Tb38-1. The immunogenic protein, and fusion proteins
XX containing one or more of the proteins or one of the proteins plus
XX ESAT-6, are useful in vaccines, preferably when formulated with a
XX non-specific adjuvant, to induce an immune response against
XX M.tuberculosis (for treatment or prevention).
XX
XX Sequence 95 AA;
XX
XX
XX Query Match 95.2%; Score 79; DB 18; Length 95;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GSLOGQWRGAAGTAA 15
XX | | | | | | | | | | | | | | |
XX Db 32 gslggqwrngaagtaa 46
XX
XX
XX RESULT 12
XX AAW32376
XX ID AAW32376 standard; Protein; 95 AA.
XX
XX AC AAW32376;
XX
XX 13-JAN-1998 (first entry)
XX
XX Mycobacterium tuberculosis antigen Tb38-1.
XX
XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9709429-A2.
XX
XX 13-MAR-1997.
XX
XX

```

```

PF 30-AUG-1996; 96WO-US14675.
XX
XX 12-JUL-1996; 96US-0680573.
XX
XX 01-SEP-1995; 95US-0523435.
XX
XX 22-SEP-1995; 95US-0532136.
XX
XX 22-MAR-1996; 96US-0620280.
XX
XX 05-JUN-1996; 96US-0658800.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky VAW;
PI Twardzik DR, Vedvick TH;
XX
XX WPI; 1997-192904/17.
XX
XX N-PSDB; AAT91445.
XX
XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
XX - useful for diagnosis of M. tuberculosis infection
XX
XX Example 3; Page 136; 190pp; English.
XX
XX A new immunogenic polypeptide has been developed comprising an
XX immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
XX its variant differing only in conservative substitutions and/or
XX modifications). The present sequence represents a M.tuberculosis
XX antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose
XX M.tuberculosis infection by forming complexes with specific
XX antibodies in the sample. Fragments of DNA encoding the immunogenic
XX polypeptide can be used as diagnostic primers or probes and agents
XX that bind to the antigen, especially monoclonal antibodies or
XX equivalent polyclonal antibodies, are also used for diagnosis.
XX
XX Sequence 95 AA;
XX
XX
XX Query Match 95.2%; Score 79; DB 18; Length 95;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GSLOGQWRGAAGTAA 15
XX | | | | | | | | | | | | | | |
XX Db 32 gslggqwrngaagtaa 46
XX
XX
XX RESULT 13
XX AAW81747
XX ID AAW81747 standard; Protein; 95 AA.
XX
XX AC AAW81747;
XX
XX 27-JAN-1999 (first entry)
XX
XX M. tuberculosis immunogenic polypeptide Tb38-1.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US18293.
XX
XX 13-MAR-1997; 97US-0818112.
XX
XX 11-OCT-1996; 96US-0730510.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky VAW, Twardzik DR, Vedvick TS;
XX

```

DR WPI; 1998-261042/23.  
 DR N-PSDB; AAV64491.  
 XX  
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX  
 PS Example 3b; Page 117; 230pp; English.  
 XX  
 CC This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (Mt) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This sequence  
 CC can be formulated into vaccines and/or pharmaceutical compositions for  
 CC immunising against M. tuberculosis infection or may be used for the  
 CC diagnosis of tuberculosis.  
 XX  
 XX Sequence 95 AA;  
 SQ

Query Match 95.2%; Score 79; DB 19; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIQGGWRGAAGTAA 15  
 |||||  
 Db 32 gslggqrgaagtaa 46

RESULT 14  
 AAW64321  
 ID AAW64321 standard; Peptide; 95 AA.  
 XX  
 AC AAW64321;  
 XX  
 DT 09-NOV-1998 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen Tb38-1 peptide.  
 XX  
 KW Tuberculosis; infection; diagnosis; antigen; Tb38-1.  
 XX  
 OS Mycobacterium tuberculosis strain H37Rv.  
 XX  
 PN WO9816645-A2.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 07-OCT-1997; 97WO-US18214.  
 XX  
 PR 13-MAR-1997; 97US-081811.  
 PR 11-OCT-1996; 96US-0729622.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1998-251292/22.  
 DR N-PSDB; AAV44384.  
 XX  
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and diagnosis of tuberculosis  
 XX  
 PS Example 3; Page 123; 250pp; English.  
 XX  
 CC This is an antigenic portion of Mycobacterium tuberculosis antigen  
 CC Tb38-1. A DNA sequence (see AAV44384) coding for antigen Tb38-1 was  
 CC isolated from a M. tuberculosis strain H37Rv expression library  
 CC using sera from patients having pulmonary or pleural tuberculosis.  
 CC The invention relates to compositions and methods for diagnosing  
 CC tuberculosis. It provides polypeptides (see AAW64291-W64379)  
 CC comprising an antigenic portion of a soluble M. tuberculosis  
 CC antigen, or an immunogenic portion of an M. tuberculosis antigen,

CC as well as DNA sequences encoding such polypeptides, recombinant  
 CC expression vectors and transformed or transfected host cells. Also  
 CC claimed are methods and diagnostic kits for detecting M.  
 CC tuberculosis infection in a patient using these polypeptides,  
 CC antibodies or oligonucleotide probes and primers, for the diagnosis  
 CC of tuberculosis.  
 XX  
 XX Sequence 95 AA;  
 SQ

Query Match 95.2%; Score 79; DB 19; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIQGGWRGAAGTAA 15  
 |||||  
 Db 32 gslggqrgaagtaa 46

RESULT 15  
 AAY32097  
 ID AAY32097 standard; Protein; 95 AA.  
 XX  
 AC AAY32097;  
 XX  
 DT 17-JAN-2000 (first entry)  
 XX  
 DE Mycobacterium tuberculosis antigen Tb38-1.  
 XX  
 KW Tuberculosis; antigen; fusion protein; Tb38-1; diagnosis; therapy;  
 KW vaccine; immunogen.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9951748-A2.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 07-APR-1999; 99WO-US07717.  
 XX  
 PR 07-APR-1998; 98US-0056556.  
 PR 30-DEC-1998; 98US-0223040.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Alderson M, Campos-Neto A;  
 XX  
 DR WPI; 1999-601610/51.  
 XX  
 PT New fusion proteins useful for diagnosis, prevention and treatment of  
 PT tuberculosis -  
 XX  
 PS Claim 1; Fig 4D; 83pp; English.  
 XX  
 CC This sequence represents the Mycobacterium tuberculosis antigen  
 CC Tb38-1. The invention provides fusion proteins (see AAY32059-71)  
 CC containing at least 2 M. tuberculosis antigens such as Tb38-1, e.g.  
 CC TbF-2 (see AAY32063) and a TbH9-Tb38-1 fusion. The new fusion  
 CC proteins are useful as vaccines for preventing tuberculosis  
 CC (claimed), for diagnosis (via in vitro assays or intradermal skin  
 CC tests for detection of anti-M. tuberculosis antibodies), monitoring  
 CC of disease progression, and treatment of tuberculosis. They are  
 CC more effective immunogens than mixtures of the individual protein  
 CC components.  
 XX  
 XX Sequence 95 AA;  
 SQ

Query Match 95.2%; Score 79; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSIQGGWRGAAGTAA 15

Db 32 gs1ggwrgaagtaa 46

Search completed: July 5, 2001, 11:45:43  
Job time: 272 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:47:00 ; Search time 61.79 Seconds  
(without alignments)  
5.216 Million cell updates/sec

Title: US-09-462-480-10  
Perfect score: 83  
Sequence: 1 GSLQGWGGAAGTAAA 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2.6/ptodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2.6/ptodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2.6/ptodata/2/1aa/6A\_COMB.pep: \*  
4: /cgn2.6/ptodata/2/1aa/6B\_COMB.pep: \*  
5: /cgn2.6/ptodata/2/1aa/PCTUS\_COMB.pep: \*  
6: /cgn2.6/ptodata/2/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	50.6	31	1 US-08-425-069-46	Sequence 46, Appl
2	42	50.6	31	2 US-08-317-844B-46	Sequence 46, Appl
3	42	50.6	34	1 US-08-425-069-42	Sequence 42, Appl
4	42	50.6	34	1 US-08-425-069-54	Sequence 54, Appl
5	42	50.6	34	2 US-08-317-844B-42	Sequence 42, Appl
6	42	50.6	34	2 US-08-317-844B-54	Sequence 54, Appl
7	42	50.6	367	2 US-08-515-251A-4	Sequence 4, Appl
8	42	50.6	718	1 US-08-425-069-2	Sequence 2, Appl
9	42	50.6	718	2 US-08-317-844B-2	Sequence 2, Appl
10	42	50.6	747	4 US-09-034-177-3	Sequence 3, Appl
11	40	48.2	1911	1 US-08-348-006B-5	Sequence 5, Appl
12	40	48.2	1911	2 US-08-800-825A-5	Sequence 5, Appl
13	40	48.2	1911	4 US-09-158-657-5	Sequence 5, Appl
14	40	48.2	1911	5 PCT-US94-10166-5	Sequence 5, Appl
15	39	47.0	759	2 US-08-450-351-2	Sequence 2, Appl
16	39	47.0	759	2 US-08-450-351-4	Sequence 4, Appl
17	39	47.0	980	1 US-08-220-151-4	Sequence 5, Appl
18	39	47.0	980	1 US-08-413-118-5	Sequence 5, Appl
19	39	47.0	980	3 US-08-473-446-5	Sequence 5, Appl
20	38.5	46.4	303	2 US-08-846-762-20	Sequence 20, Appl
21	38.5	46.4	303	2 US-08-846-762-91	Sequence 91, Appl
22	38.5	46.4	441	2 US-08-491-835-4	Sequence 4, Appl
23	38.5	46.4	441	3 US-08-946-092A-4	Sequence 4, Appl
24	38.5	46.4	441	4 US-09-172-062-4	Sequence 4, Appl
25	38.5	46.4	441	5 PCT-US94-00685-4	Sequence 4, Appl
26	38	45.8	30	1 US-08-425-069-58	Sequence 58, Appl
27	38	45.8	30	2 US-08-317-844B-58	Sequence 58, Appl

28	38	45.8	34	1 US-08-425-069-52	Sequence 52, Appl
29	38	45.8	34	2 US-08-317-844B-52	Sequence 52, Appl
30	38	45.8	261	2 US-07-857-224B-58	Sequence 58, Appl
31	38	45.8	284	2 US-08-320-148B-2	Sequence 2, Appl
32	38	45.8	284	3 US-08-589-028-6	Sequence 6, Appl
33	38	45.8	284	4 US-08-784-582-6	Sequence 6, Appl
34	38	45.8	284	4 US-08-785-271-6	Sequence 2, Appl
35	38	45.8	284	4 US-09-031-898-2	Sequence 2, Appl
36	38	45.8	345	2 US-08-332-562A-132	Sequence 132, App
37	38	45.8	588	1 US-08-391-615-5	Sequence 5, Appl
38	38	45.8	756	1 US-08-434-730-16	Sequence 16, Appl
39	37	44.6	253	2 US-08-659-251-4	Sequence 4, Appl
40	37	44.6	253	4 US-09-256-490-4	Sequence 4, Appl
41	37	44.6	253	5 PCT-US96-11445-4	Sequence 4, Appl
42	37	44.6	260	3 US-08-906-769-139	Sequence 139, App
43	37	44.6	260	4 US-08-906-616-139	Sequence 139, App
44	37	44.6	260	4 US-08-639-075A-139	Sequence 139, App
45	37	44.6	260	4 US-09-012-431-139	Sequence 139, App

ALIGNMENTS

RESULT 1  
US-08-425-069-46  
; Sequence 46, Application US/08425069  
; Patent No. 5728810  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; APPLICANT: Hinman, Michael B.  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5728810th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/425,069  
; FILING DATE: 19-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-106P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 31 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: nephila clavipes  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..31

OTHER INFORMATION: /label= silkl\_repeat  
US-08-425-069-46

Query Match 50.6%; Score 42; DB 1; Length 31;  
Best Local Similarity 62.5%; Pred. No. 2.2;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAAA 16  
| | | | |  
Db 4 GGLGGQGAGAAAAA 19

## RESULT 2

US-08-317-844B-46  
; Sequence 46, Application US/08317844B  
; Patent No. 5989894  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; APPLICANT: Himman, Michael B.  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5989894th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,844B  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-105P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300  
; TELEFAX: (703) 241-2848  
; TELEX: 248345

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: nephila clavipes  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..31  
OTHER INFORMATION: /label= silkl\_repeat  
US-08-317-844B-46

Query Match 50.6%; Score 42; DB 2; Length 31;  
Best Local Similarity 62.5%; Pred. No. 2.2;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAAA 16  
| | | | |

Db 4 GGLGGQGAGAAAAA 19

## RESULT 3

US-08-425-069-42  
; Sequence 42, Application US/08425069  
; Patent No. 5728810  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; APPLICANT: Himman, Michael B.  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5728810th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/425,069  
; FILING DATE: 19-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-106P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; TELEX:

INFORMATION FOR SEQ ID NO: 42:

SEQUENCE CHARACTERISTICS:  
LENGTH: 34 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: nephila clavipes  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..34  
OTHER INFORMATION: /label= silkl\_repeat  
US-08-425-069-42

Query Match 50.6%; Score 42; DB 1; Length 34;  
Best Local Similarity 62.5%; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAAA 16  
| | | | |  
Db 4 GGLGGQGAGAAAAA 19

## RESULT 4

US-08-425-069-54  
; Sequence 54, Application US/08425069  
; Patent No. 5728810  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming



APPLICANT: Hinman, Michael B.  
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 301 No. 5728810th Washington Street  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,069  
FILING DATE: 19-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX:  
INFORMATION FOR SEQ ID NO: 54:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: nephila clavipes  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..34  
OTHER INFORMATION: /label= silkl\_repeat  
US-08-425-069-54

Query Match 50.6%; Score 42; DB 1; Length 34;  
Best Local Similarity 62.5%; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLQGQWRGAAGTAA 16  
| | | | | | | | | |  
Db 4 GGLGGGAGAGAAAAA 19

RESULT 5  
US-08-317-844B-42  
Sequence 42, Application US/08317844B  
Patent No. 5989894  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Xu, Ming  
APPLICANT: Hinman, Michael B.  
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 301 No. 5989894th Washington Street  
CITY: Falls Church  
STATE: Virginia

COUNTRY: U.S.A.  
ZIP: 22046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,844B  
FILING DATE: 04-OCT-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-105P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 241-1300  
TELEFAX: (703) 241-2848  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: nephila clavipes  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..34  
OTHER INFORMATION: /label= silkl\_repeat  
US-08-317-844B-42

Query Match 50.6%; Score 42; DB 2; Length 34;  
Best Local Similarity 62.5%; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLQGQWRGAAGTAA 16  
| | | | | | | | | |  
Db 4 GGLGGGAGAGAAAAA 19

RESULT 6  
US-08-317-844B-54  
Sequence 54, Application US/08317844B  
Patent No. 5989894  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Xu, Ming  
APPLICANT: Hinman, Michael B.  
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 301 No. 5989894th Washington Street  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,844B  
FILING DATE: 04-OCT-1994

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-105P  
TELEPHONE: (703) 241-1300  
TELEFAX: (703) 241-2848  
TELEX: 248345  
INFORMATION FOR SEQ ID NO: 54:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 34 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: nephila clavipes  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1-34  
OTHER INFORMATION: /label= silkl\_repeat  
US-08-317-844B-54

Query Match 50.6%; Score 42; DB 2; Length 34;  
Best Local Similarity 62.5%; Pred. No. 2.4;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLQGWGRGAAGTAA 16  
DB 4 GGLGGGAGAAAAAA 19

RESULT 7  
US-08-515-251A-4  
Sequence 4, Application US/08515251A  
Patent No. 5891677  
GENERAL INFORMATION:  
APPLICANT: GERLACH, GERALD F.  
APPLICANT: WILLSON, PHILIP J.  
APPLICANT: ROSSI-CAMPOS, AMALIA  
APPLICANT: POTTER, ANDREW A.  
TITLE OF INVENTION: MEMBRANE LIPOPROTEIN A AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ROBINS & ASSOCIATES  
STREET: 90 MIDDLEFIELD ROAD, SUITE 200  
CITY: MENLO PARK  
STATE: CA  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/515,251A  
FILING DATE: 15-AUG-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,558  
FILING DATE: 05-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINS, ROBERTA L.  
REGISTRATION NUMBER: 33,208  
REFERENCE/DOCKET NUMBER: 9001-0027.10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 325-7812

TELEFAX: (650) 325-7823  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 367 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-515-251A-4

Query Match 50.6%; Score 42; DB 2; Length 367;  
Best Local Similarity 42.9%; Pred. No. 32;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 SLOGQWGRGAAGTAA 15  
DB 347 SIKGQGVIGATA 360

RESULT 8  
US-08-425-069-2  
Sequence 2, Application US/08425069  
Patent No. 5728810  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Xu, Ming  
APPLICANT: Hinman, Michael B.  
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: 301 No. 5728810th Washington Street  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,069  
FILING DATE: 19-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 718 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-425-069-2

Query Match 50.6%; Score 42; DB 1; Length 718;  
Best Local Similarity 62.5%; Pred. No. 67;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLQGWGRGAAGTAA 16  
DB 100 GGLGGGAGAAAAAA 115

RESULT 9  
US-08-317-844B-2  
; Sequence 2, Application US/08317844B  
; Patent No. 5989894  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Xu, Ming  
; APPLICANT: Hinman, Michael B.  
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: 301 No. 5989894th Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22046  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/317,844B  
; FILING DATE: 04-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-105P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 241-1300  
; TELEFAX: (703) 241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 718 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-317-844B-2

Query Match 50.6%; Score 42; DB 2; Length 718;  
Best Local Similarity 62.5%; Pred. No. 67;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLQGGQWRGAAGTAAA 16  
| | | | |  
Db 100 GGLGGGAGAGAAAAA 115

RESULT 10  
US-09-034-177-3  
; Sequence 3, Application US/09034177  
; Patent No. 6127145  
; GENERAL INFORMATION:  
; APPLICANT: Lal, Preeti  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette

Query Match 50.6%; Score 42; DB 2; Length 718;  
Best Local Similarity 62.5%; Pred. No. 67;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLQGGQWRGAAGTAAA 16  
| | | | |  
Db 100 GGLGGGAGAGAAAAA 115

RESULT 11  
US-08-348-006B-5  
; Sequence 5, Application US/08348006B  
; Patent No. 5658756  
; GENERAL INFORMATION:  
; APPLICANT: RODAN, GIDEON A.  
; APPLICANT: SCHMIDT, AZRIEL  
; APPLICANT: RUTLEDGE, SU JANE  
; TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASE  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: J. MARK HAND  
; STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
; CITY: RAHWAY  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07065-0900  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/348,006B  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/122,032  
; FILING DATE: 14-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAND, J., MARK  
; REGISTRATION NUMBER: 36,545  
; REFERENCE/DOCKET NUMBER: 189921A

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/034,177  
FILING DATE: HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0486 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 747 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: GI 1174414  
US-09-034-177-3

Query Match 50.6%; Score 42; DB 4; Length 747;  
Best Local Similarity 62.5%; Pred. No. 70;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLQGGQWRGAAGTAAA 16  
| | | | |  
Db 100 GGLGGGAGAGAAAAA 115

RESULT 11  
US-08-348-006B-5  
; Sequence 5, Application US/08348006B  
; Patent No. 5658756  
; GENERAL INFORMATION:  
; APPLICANT: RODAN, GIDEON A.  
; APPLICANT: SCHMIDT, AZRIEL  
; APPLICANT: RUTLEDGE, SU JANE  
; TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASE  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: J. MARK HAND  
; STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
; CITY: RAHWAY  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07065-0900  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/348,006B  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/122,032  
; FILING DATE: 14-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAND, J., MARK  
; REGISTRATION NUMBER: 36,545  
; REFERENCE/DOCKET NUMBER: 189921A

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-3905  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1911 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-348-006B-5

Query Match 48.2%; Score 40; DB 1; Length 1911;  
Best Local Similarity 57.1%; Pred. No. 3.9e+02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTA 14  
||| :| |||  
Db 825 GSLARWEPPAGTA 838

## RESULT 12

US-08-800-825A-5  
Sequence 5, Application US/08800825A  
Patent No. 5866397  
GENERAL INFORMATION:  
APPLICANT: RODAN, GIDEON A.  
APPLICANT: SCHMIDT, AZRIEL  
TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASE  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: J. MARK HAND - MERCK & CO., INC.  
STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,825A  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: HAND, J. MARK  
REGISTRATION NUMBER: 36,545  
REFERENCE/DOCKET NUMBER: 18992DA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732-594-3905  
TELEFAX: 732-594-4720  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1911 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-800-825A-5

Query Match 48.2%; Score 40; DB 2; Length 1911;  
Best Local Similarity 57.1%; Pred. No. 3.9e+02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTA 14  
||| :| |||  
Db 825 GSLARWEPPAGTA 838

## RESULT 13

US-09-158-657-5  
Sequence 5, Application US/09158657  
Patent No. 6214564  
GENERAL INFORMATION:  
APPLICANT: RODAN, GIDEON A.  
APPLICANT: SCHMIDT, AZRIEL  
TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASE  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: J. MARK HAND - MERCK & CO., INC.  
STREET: 126 E. LINCOLN AVE., P.O. BOX 2000  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/158,657  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/800,825  
FILING DATE: 14-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: HAND, J. MARK  
REGISTRATION NUMBER: 36,545  
REFERENCE/DOCKET NUMBER: 18992DA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 732-594-3905  
TELEFAX: 732-594-4720  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1911 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-158-657-5

Query Match 48.2%; Score 40; DB 4; Length 1911;  
Best Local Similarity 57.1%; Pred. No. 3.9e+02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTA 14  
||| :| |||  
Db 825 GSLARWEPPAGTA 838

## RESULT 14

PCT-US94-10166-5  
Sequence 5, Application PC/TUS9410166  
GENERAL INFORMATION:  
APPLICANT: RODAN, GIDEON A.  
APPLICANT: SCHMIDT, AZRIEL  
APPLICANT: RUTLEDGE, SU JANE  
TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JOHN W. WALLEN III  
STREET: P.O. BOX 2000, 126 E. LINCOLN AVE.  
CITY: RAHWAY  
STATE: NJ

COUNTRY: USA  
 ZIP: 07065  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/10166  
 FILING DATE: 09-SEPT-1994  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/122,032  
 FILING DATE: 14-SEP-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: WALLEN, JOHN W III  
 REGISTRATION NUMBER: 35403  
 REFERENCE/DOCKET NUMBER: 18992  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 908-594-3905  
 TELEFAX: 908-594-4720  
 TELEX: 138825  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1911 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US94-10166-5

Query Match 48.2%; Score 40; DB 5; Length 1911;  
 Best Local Similarity 57.1%; Pred. No. 3.9e+02;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 GSLQGWGGAAGTA 14  
 III :I II  
 DB 825 GSLARWEPAGTA 838

RESULT 15  
 US-08-450-351-2  
 Sequence 2, Application US/08450351  
 Patent No. 5981213  
 GENERAL INFORMATION:  
 APPLICANT: Hansen, Eric J.  
 APPLICANT: Helminen, Merja E.  
 APPLICANT: Maciver, Isobel  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Arnold, White & Durkee  
 STREET: P.O. Box 4433  
 CITY: Houston  
 STATE: Texas  
 COUNTRY: USA  
 ZIP: 77210  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/450,351  
 FILING DATE: 25-MAY-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Highlander, Steven L.  
 REGISTRATION NUMBER: 37,642  
 REFERENCE/DOCKET NUMBER: AMCY:019  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000  
 TELEFAX: (512) 474-7577  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 759 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 US-08-450-351-2

Query Match 47.0%; Score 39; DB 2; Length 759;  
 Best Local Similarity 50.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSLQGWGGAAGTA 14  
 I :I I I I  
 DB 613 GYTEGAWRARAGVA 626

Search completed: July 5, 2001, 11:47:01  
 Job time: 315 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:48:35 ; Search time 79.63 seconds  
(without alignments)  
15.306 Million cell updates/sec

Title: US-09-462-480-10  
Perfect score: 83  
Sequence: 1 GSLOGQWRGAAGTAA 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	95.2	100	2 H70802	hypothetical prote
2	48	57.8	103	2 B70600	hypothetical prote
3	44	53.0	666	2 G70695	hypothetical prote
4	44	53.0	1186	2 T35661	probable chromosom
5	43	51.8	124	2 T10919	3C3.10 protein - S
6	43	51.8	203	2 JC6113	melanocyte-specifi
7	43	51.8	255	2 D64833	probable ABC-type
8	43	51.8	255	2 F85619	hypothetical prote
9	43	51.8	450	2 C75033	adenylosuccinate 1
10	43	51.8	450	2 H71135	probable adenylosu
11	43	51.8	542	2 A70826	probable fadE8 pro
12	42	50.6	179	2 A85217	hypothetical prote
13	42	50.6	277	2 T04441	hypothetical prote
14	42	50.6	367	2 I36649	lipoprotein - Acti
15	42	50.6	506	2 A40679	transcription enha
16	42	50.6	523	2 B40679	transcription enha
17	42	50.6	533	2 T03441	probable ammonium
18	42	50.6	718	2 A36068	major ampullate fi
19	41	49.4	69	2 D49786	lactococcin A1 pre
20	41	49.4	294	2 D70525	probable beta-1 -
21	41	49.4	315	2 S23108	methionyl-tRNA for
22	41	49.4	315	2 E85993	hypothetical prote
23	41	49.4	361	2 T30743	hypothetical prote
24	41	49.4	381	2 B83450	conserved hypotet
25	41	49.4	427	2 E72488	probable tryptopa
26	41	49.4	528	2 T02863	probable membrane
27	41	49.4	1050	3 JC7578	endo-1,4-beta-xyla
28	40.5	48.8	389	2 JC4001	macrolide 3-O-acyl
29	40.5	48.8	1443	2 G75393	hypothetical prote

ALIGNMENTS

RESULT 1

H70802  
hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70802  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A: Reference number: A70500; MUID: 98295987  
A: Accession: H70802  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-100 <COL>  
A: Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAI7966.1; PID:g296  
A: Experimental source: strain H37RV  
C: Genetics:  
A: Gene: RV3874

Query Match 95.2% Score 79; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 2.3e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 GSLOGQWRGAAGTAA 15  
|||||  
DB 37 GSLOGQWRGAAGTAA 51

RESULT 2

B70600  
hypothetical protein RV3905c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: B70600  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A: Reference number: A70500; MUID: 98295987  
A: Accession: B70600  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-103 <COL>  
A: Cross-references: GB:294121; GB:AL123456; NID:g3261736; PIDN:CAB08096.1; PID:e3127  
A: Experimental source: strain H37RV  
C: Genetics:

A:Gene: RV3905c

Query Match 57.8%; Score 48; DB 2; Length 103;  
Best Local Similarity 57.1%; Pred. No. 1.4;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTA 14  
| : | | | | | : |  
Db 40 GQMLGGWRGASGSA 53

RESULT 3

G70695 hypothetical protein RV3779 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: G70695

R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, C.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: G70695

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-666 <COL>

A:Cross-references: GB:280343; GB:AL123456; NID:g3261648; PIDN:CAB02458.1; PID:g1552862

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3779

C:Superfamily: Mycobacterium tuberculosis hypothetical protein RV3779

Query Match 53.0%; Score 44; DB 2; Length 666;  
Best Local Similarity 76.9%; Pred. No. 34;  
Matches 10; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

Qy 6 QWR--GAAGTAA 16  
: | | | | | | |  
Db 231 EWRAAGAACTAA 243

RESULT 4

T35661

probable chromosome associated protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 02-Jun-2000  
C:Accession: T35661  
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, December 1998

A:Reference number: Z21585

A:Accession: T35661

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1186 <MUR>

A:Cross-references: EMBL:AL034447; PIDN:CAA22420.1; GSPDB:GN00070; SCOEDB:SC7A1.21

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC7A1.21

C:Superfamily: chromosome segregation protein SMC1

Query Match 53.0%; Score 44; DB 2; Length 1186;  
Best Local Similarity 71.4%; Pred. No. 60;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTA 14  
| : | | | | | : |  
Db 718 GRLAQARGAAGEA 731

RESULT 5

T10919

3C3.10 protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C:Accession: T10919

R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1998

A:Reference number: Z17215

A:Accession: T10919

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-124 <PAR>

A:Cross-references: EMBL:AL031231; NID:el315070; PID:el315080

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: 3C3.10

Query Match 51.8%; Score 43; DB 2; Length 124;  
Best Local Similarity 70.0%; Pred. No. 9.8;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 LQGWGGAAG 12  
: | | | | |  
Db 52 IQGGWGAAG 61

RESULT 6

JC6113

melanocyte-specific protein 1 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 23-Mar-1997 #sequence\_revision 09-May-1997 #text\_change 05-Nov-1999

C:Accession: JC6113

R:Shioda, T.; Fenner, M.H.; Isselbacher, K.J.

Proc. Natl. Acad. Sci. U.S.A. 93, 12298-12303, 1996

A:Title: msq1, a novel melanocyte-specific gene, encodes a nuclear protein and is ess

A:Reference number: JC6113; MUID:97057236

A:Accession: JC6113

A:Molecule type: mRNA

A:Residues: 1-203 <SHI>

A:Cross-references: GB:U65091; NID:g1854000; PIDN:AAC53048.1; PID:g1854001

C:Comment: This protein is a nuclear protein whose expression is confined to pigment

C:Genetics:

A:Gene: msq1

Query Match 51.8%; Score 43; DB 2; Length 203;  
Best Local Similarity 64.3%; Pred. No. 16;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0

Qy 3 LQGWGGAAGTAA 16  
| : | | | | | | |  
Db 102 LNSQYGAATATA 115

RESULT 7

D64833

probable ABC-type transport protein ycbE - Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 02-Feb-2001

C:Accession: D64833

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A:Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: D64833

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-255 <BLAT>

A:Cross-references: GB:AE000195; GB:U00096; NID:g1787156; PIDN:AAC74019.1; PID:g17871



A: Experimental source: strain K-12, substrain MG1655

C: Gene: ycbE

C: Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C: Keywords: ATP; nucleotide binding; P-loop; transport protein

F: 27-209/Domain: ATP-binding cassette homology <ABC>

F: 44-51/Region: nucleotide-binding motif A (P-loop)

Query Match 51.8%; Score 43; DB 2; Length 255;

Best Local Similarity 64.3%; Pred. No. 20;

Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 LQOWRGAAGTAA 16

I: I I I I I I I

Db 105 LKQWRDAARRALA 118

RESULT 8

F85619

hypothetical protein ycbE [imported] - Escherichia coli (strain O157:H7)

C: Species: Escherichia coli

C: Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001

C: Accession: F85619

R: Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001

A: Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A: Reference number: A85480; MUID: 21074935; PMID: 11206551

A: Accession: F85619

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-255 <STO>

A: Cross-references: GB:AE005174; NID: g12514108; PIDN: AAC55418.1; GSPDB: GN00145; UWGP: 212

A: Experimental source: strain O157:H7, substrain EDL933

C: Genetics:

A: Gene: ycbE

Query Match 51.8%; Score 43; DB 2; Length 255;

Best Local Similarity 64.3%; Pred. No. 20;

Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 LQOWRGAAGTAA 16

I: I I I I I I I

Db 105 LKQWRDAARRALA 118

RESULT 9

C75033

adenylosuccinate lyase (purb) PAB0829 - Pyrococcus abyssi (strain Orsay)

C: Species: Pyrococcus abyssi

C: Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C: Accession: C75033

R: Anonymous, Genoscope

A: Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A: Reference number: A75001

A: Accession: C75033

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-450 <RAW>

A: Cross-references: GB:AJ248287; GB:AL096836; NID: g5458657; PIDN: CAB50160.1; PID: g545866

A: Experimental source: strain Orsay

C: Genetics:

A: Gene: purB; PAB0829

C: Superfamily: fumarate hydratase

Query Match 51.8%; Score 43; DB 2; Length 450;

Best Local Similarity 64.3%; Pred. No. 34;

Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LQOWRGAAGTAA 16

I: I I I I I I I

Db 184 LVKMRGAVGTAA 197

RESULT 10

H71135

probable adenylosuccinate lyase - Pyrococcus horikoshii

C: Species: Pyrococcus horikoshii

C: Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Jul-2000

C: Accession: H71135

R: Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu  
DNA Res. 5, 55-76, 1998

A: Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A: Reference number: A71000; MUID: 98344137

A: Accession: H71135

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-450 <KAW>

A: Cross-references: GB:AP000003; NID: g3236130; PIDN: BAA29946.1; PID: g3257263

A: Experimental source: strain OT3

A: Note: This accession replaces an interim accession for a sequence replaced by GenBa

C: Genetics:

A: Gene: PH0852

C: Superfamily: fumarate hydratase

Query Match 51.8%; Score 43; DB 2; Length 450;

Best Local Similarity 64.3%; Pred. No. 34;

Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LQOWRGAAGTAA 16

I: I I I I I I I

Db 184 LVKMRGAVGTAA 197

RESULT 11

A70826

probable fadE8 protein - Mycobacterium tuberculosis (strain H37RV)

C: Species: Mycobacterium tuberculosis

C: Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C: Accession: A70826

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A: Reference number: A70500; MUID: 98295987

A: Accession: A70826

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-542 <COL>

A: Cross-references: GB:AL021943; GB:AL123456; NID: g3261530; PIDN: CAA17469.1; PID: e129

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: fadE8

Query Match 51.8%; Score 43; DB 2; Length 542;

Best Local Similarity 64.3%; Pred. No. 40;

Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 LQOWRGAAGTAA 16

I: I I I I I I I

Db 514 LGQWGGAYGTMPA 527

RESULT 12

A85217

hypothetical protein A74g19200 [imported] - Arabidopsis thaliana

C: Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: A85217  
R:anonymus, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: A85217  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-179 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7268715; PIDN:CAB78922.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g19200  
A:Map position: 4

Query Match 50.6%; Score 42; DB 2; Length 179;  
Best Local Similarity 62.5%; Pred. No. 20;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTAA 16  
I I I I I I I I I I  
Db 91 GGLGGMIAAGAAAA 106

RESULT 13  
T04441  
hypothetical protein T18B16.170 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: T04441  
R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.  
submitted to the Protein Sequence Database, April 1998  
A:Reference number: Z15359  
A:Accession: T04441  
A:Molecule type: DNA  
A:Residues: 1-277 <BEV>  
A:Cross-references: EMBL:AL021687  
A:Experimental source: cultivar Columbia; BAC clone T18B16  
C:Genetics:  
A:Map position: 4  
A:Introns: 87/1; 144/1; 179/1; 207/3  
A:Note: T18B16.170

Query Match 50.6%; Score 42; DB 2; Length 277;  
Best Local Similarity 62.5%; Pred. No. 30;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTAA 16  
I I I I I I I I I I  
Db 91 GGLGGMIAAGAAAA 106

RESULT 14  
I39649  
lipoprotein - Actinobacillus pleuropneumoniae  
C:Species: Actinobacillus pleuropneumoniae  
C>Date: 19-Jul-1996 #sequence\_revision 19-Jul-1996 #text\_change 17-Nov-2000  
C:Accession: I39649; I39518; S52740  
R:Bunka, S.; Christensen, C.; Potter, A.A.; Willson, P.J.; Gerlach, G.F.  
Infect. Immun. 63, 2797-2800, 1995  
A>Title: Cloning and characterization of a protective outer membrane lipoprotein of Actinobacillus pleuropneumoniae  
A:Reference number: I39649; MUID:95310047  
A:Accession: I39649  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-367 <RES>  
A:Cross-references: EMBL:Z48920; NID:g757528; PIDN:CAA88756.1; PID:g757529  
R:Ito, H.; Uchida, I.; Sekizaki, T.; Oishi, E.; Kawai, T.; Okabe, T.; Taneno, A.; Terak  
Microb. Pathog. 18, 29-36, 1995  
A>Title: Molecular cloning of an Actinobacillus pleuropneumoniae outer membrane lipoprotein  
A:Reference number: I39518; MUID:95302932

A:Accession: I39518  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-361, 'KEDK', 366 <RE2>  
A:Cross-references: GB:D28491; NID:g808858; PIDN:BAA05852.1; PID:g808859  
C:Genetics:  
A:Gene: omIA

Query Match 50.6%; Score 42; DB 2; Length 367;  
Best Local Similarity 42.9%; Pred. No. 39;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SLOGQWRGAAGTAA 15  
I I I I I I I I I I  
Db 347 SIKGQGVIGATA 360

RESULT 15  
A40679  
transcription enhancer regulatory element TRE-2S, splice form 1 - human T-cell lympho  
N:Alternate names: transcription holding protein 1 (THP-1)  
C:Species: human T-cell lymphotropic virus type 1, HTLV-1  
C>Date: 16-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
C:Accession: A40679  
R:Tanimura, A.; Teshima, H.; Fujisawa, J.; Yoshida, M.  
J. Virol. 67, 5375-5382, 1993  
A>Title: A new regulatory element that augments the Tax-dependent enhancer of human T  
A:Reference number: A40679; MUID:9335614  
A:Accession: A40679  
A>Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-506 <TAN>  
A:Cross-references: GB:D14827; NID:g439899; PIDN:BAA03568.1; PID:g488506  
A:Experimental source: HUT102 cells  
A:Note: sequence extracted from NCBI backbone (NCBIN:136287, NCBI:136288)  
C:Superfamily: gII transforming protein

Query Match 50.6%; Score 42; DB 2; Length 506;  
Best Local Similarity 55.0%; Pred. No. 54;  
Matches 11; Conservative 0; Mismatches 5; Indels 4; Gaps 1;

Qy 1 GSLOGQW----RGAAAGTAA 16  
I I I I I I I I I I  
Db 423 GKIQWQWGRARGAAAPAA 442

Search completed: July 5, 2001, 11:48:35  
Job time: 374 sec

Thu Jul 5 13:58:39 2001

us-09-462-480-10.rpr

Page 5

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:51:44 ; Search time 41.8 Seconds  
(without alignments)  
13.112 Million cell updates/sec

Title: US-09-462-480-10  
Perfect score: 83  
Sequence: 1 GSLQGQWRGAAGTAAA 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	44	53.0	321	1	GUN1_STRHA	P33682	streptomyc
2	43	51.8	203	1	MSG1_MOUSE	P97769	mus musc
3	43	51.8	255	1	SSUB_ECOLI	P38053	escherichia
4	43	51.8	450	1	PUR8_PYRAB	Q9uz99	pyrococcus
5	43	51.8	450	1	PUR8_PYRHO	O58582	pyrococcus
6	42	50.6	747	1	SPD1_NEPCL	P19837	nephila cila
7	41	49.4	314	1	FMT_ECOLI	P23882	escherichia
8	41	49.4	427	1	TRB1_AERPE	Q9y8t5	aeropyrum p
9	40	48.2	132	1	YJW1_ECOLI	P39394	escherichia
10	40	48.2	513	1	COX1_ORNAN	Q36452	ornithorhyn
11	40	48.2	513	1	NIFK_BRASP	P06122	bradyrhizob
12	40	48.2	514	1	COX1_HALGR	P38595	halichoerus
13	39	47.0	55	1	NXB2_CERLA	P01526	cerebratulu
14	39	47.0	55	1	NXB4_CERLA	P01525	cerebratulu
15	39	47.0	156	1	COX1_COTJA	P24984	coturnix co
16	39	47.0	250	1	YMI9_MYCTU	Q10405	mycobacteri
17	39	47.0	318	1	FMT_HAEIN	P44787	haemophilus
18	39	47.0	431	1	NFX1_CHICK	Q90932	gallus gall
19	39	47.0	444	1	PUR8_ARCFU	O28041	archaeoglob
20	39	47.0	515	1	COX1_CHICK	P18943	gallus gall
21	39	47.0	517	1	COX1_ASTPE	Q33820	asterina pe
22	39	47.0	901	1	MAUT_ECOLI	P06993	escherichia
23	39	47.0	979	1	VGLB_HSVEL	P25218	equine herp
24	39	47.0	980	1	VGLB_HSVEL	P18050	equine herp
25	39	47.0	980	1	VGLB_HSVEL	P18551	equine herp
26	39	47.0	980	1	VGLB_HSVEL	P28922	equine herp
27	38.5	46.4	441	1	GDF9_MOUSE	Q07100	mus musc
28	38	45.8	270	1	FWDC_METTH	O27605	methanobact
29	38	45.8	270	1	FWDC_METTM	O59579	methanobact
30	38	45.8	270	1	FWDC_METWO	O74031	methanobact
31	38	45.8	284	1	IPF1_MOUSE	P52946	mus musc
32	38	45.8	312	1	FMT_MYCTU	P71674	mycobacteri
33	38	45.8	381	1	IDHB_MACFA	Q28479	macaca fasc

34	38	45.8	385	1	IDHB_HUMAN	O43837	homo sapien
35	38	45.8	501	1	BACE_HUMAN	P58817	homo sapien
36	38	45.8	501	1	BACE_MOUSE	P58818	mus musc
37	38	45.8	501	1	BACE_RAT	P58819	rattus norv
38	38	45.8	518	1	NIFK_BRAJA	P20621	bradyrhizob
39	38	45.8	576	1	ACEA_ARATH	P28297	arabidopsis
40	38	45.8	590	1	SRC2_DROME	P08630	drosophila
41	38	45.8	672	1	Y442_MYCPN	P75072	mycoplasma
42	38	45.8	730	1	GELS_HORSE	Q28372	equus cabal
43	38	45.8	1070	1	Y355_HUMAN	O15063	homo sapien
44	37.5	45.2	420	1	PHR_THETH	P37250	thermus aqu
45	37	44.6	253	1	NEF_HV2KR	Q74127	human immun

## ALIGNMENTS

RESULT 1  
GUN1\_STRHA  
ID GUN1\_STRHA STANDARD; PRT; 321 AA.  
AC P33682;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE ENDOGLUCANASE 1 PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE 1)  
DE (CELLULOSE I) (CMCASE I) (CEL1).  
GN CELAL.  
OS Streptomyces halstedii.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=1944;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-33.  
RC STRAIN-JM8 / CECT3310;  
RX MEDLINE=93015685; PubMed=1400190;  
RA Fernandez-Abalos J.M., Sanchez P., Coll-Presno P.M.,  
RA Villanueva J.R., Perez P., Santamaria R.I.;  
RT "Cloning and nucleotide sequence of celal, and  
RT endo-beta-1,4-glucanase-encoding gene from Streptomyces halstedii  
JM8".  
RT J. Bacteriol. 174:6368-6376(1992).  
CC -!- FUNCTION: IMPLICATED IN THE MECHANISM OF INDUCTION EXERTED BY  
CC CELLOBIOSE.  
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
CC LINKAGES IN CELLULOSE.  
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL  
CC HYDROLASES).  
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; Z12157; CAA78145.1; -.  
CC HSSP; P26222; ITML.  
DR InterPro; IPR001524; -.  
DR Pfam; PF01341; Glyco\_hydro\_6; 1.  
DR PRINTS; PR00733; GLHYDRLASE6.  
DR PROSITE; PS00655; GLYCOSYL\_HYDROL\_F6\_1; FALSE\_NEG.  
DR PROSITE; PS00856; GLYCOSYL\_HYDROL\_F6\_2; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 27  
FT CHAIN 28 321 ENDOGLUCANASE 1.  
FT ACT\_SITE 110 110 BY SIMILARITY.  
FT ACT\_SITE 149 149 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 295 295 NUCLEOPHILE (BY SIMILARITY).  
FT DISULFID 112 156 BY SIMILARITY.  
SQ SEQUENCE 321 AA; 33694 MW; DC5E39EF3CCD114C CRC64;

modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL AE000195; AAC74019.1; --  
CC DR EMBL D90731; BAA35685.1; --  
CC DR EMBL D90732; BAA35688.1; --  
CC DR EMBL AE237695; CAB40393.1; --  
CC DR EMBL M15273; -- NOT\_ANNOTATED\_CDS.  
CC HSSP: P13569; lNBD.  
CC EcoGene: EG12358; ssuB.  
CC InterPro: IPR001617; --  
CC Pfam: PF00005; Arg\_trop.1

```
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport; Inner membrane.
FT NP_BIND 44 51 ATP (POTENTIAL).
FT CONFLICT 161 165 GALDA -> RGAGR (IN REF. 4).
FT CONFLICT 232 233 EL -> DV (IN REF. 4).
FT CONFLICT 236 237 EV -> RS (IN REF. 4).
SQ SEQUENCE 255 AA; 27738 MW; 6FA276AD6E7FACE1 CRC64;

Query Match 51.8%; Score 43; DB 1; Length 255;
Best Local Similarity 64.3%; Pred. No. 9.3;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 LQGWGGAAGTAA 16
   |:||| | |
Db 105 LKGWRDARRALA 118

RESULT 4
PUR8_PYRAB STANDARD; PRT; 450 AA.
AC OSU259;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADENYLOSUCCLINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCLINASE) (ASL).
GN PURB OR PAB0829.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
   structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 1-(5'-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)
   -5-AMINOIMIDAZOLE = FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-
   IMIDAZOLECARBOXAMIDE (ALSO CATALYZES: N6-(1,2-DICARBOXYETHYL)AMP =
   FUMARATE + AMP).
CC -1- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE LYASE 1 FAMILY. ADENYLOSUCCLINATE LYASE
   SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC EMBL: AJ248287; CAB50160.1;
DR PROSITE: PS00163; FUMARATE_LYASES; 1.
KW Purine biosynthesis; Lyase.
FT ACT_SITE 76 76 ACID (BY SIMILARITY).
FT ACT_SITE 149 149 BASE (BY SIMILARITY).
SQ SEQUENCE 450 AA; 51455 MW; 82A3C0131455CDBD CRC64;

Query Match 51.8%; Score 43; DB 1; Length 450;
Best Local Similarity 64.3%; Pred. No. 16;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LQGWGGAAGTAA 16
   |:||| | |
Db 184 LVGKMGAVGTAAS 197

RESULT 5
PUR8_PYRHO STANDARD; PRT; 450 AA.
AC OSU259;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADENYLOSUCCLINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCLINASE) (ASL).
GN PURB OR PAB0829.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
   structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 1-(5'-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)
   -5-AMINOIMIDAZOLE = FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-
   IMIDAZOLECARBOXAMIDE (ALSO CATALYZES: N6-(1,2-DICARBOXYETHYL)AMP =
   FUMARATE + AMP).
CC -1- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE LYASE 1 FAMILY. ADENYLOSUCCLINATE LYASE
   SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC EMBL: AJ248287; CAB50160.1;
DR PROSITE: PS00163; FUMARATE_LYASES; 1.
KW Purine biosynthesis; Lyase.
FT ACT_SITE 76 76 ACID (BY SIMILARITY).
FT ACT_SITE 149 149 BASE (BY SIMILARITY).
SQ SEQUENCE 450 AA; 51455 MW; 82A3C0131455CDBD CRC64;

Query Match 51.8%; Score 43; DB 1; Length 450;
Best Local Similarity 64.3%; Pred. No. 16;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LQGWGGAAGTAA 16
   |:||| | |
Db 184 LVGKMGAVGTAAS 197

RESULT 6
SPDL_NEPCL STANDARD; PRT; 747 AA.
AC PI9837;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
```

```
AC OS582;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADENYLOSUCCLINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCLINASE) (ASL).
GN PURB OR PH0852.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawanabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
   thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: 1-(5'-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)
   -5-AMINOIMIDAZOLE = FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-
   IMIDAZOLECARBOXAMIDE (ALSO CATALYZES: N6-(1,2-DICARBOXYETHYL)AMP =
   FUMARATE + AMP).
CC -1- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE LYASE 1 FAMILY. ADENYLOSUCCLINATE LYASE
   SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC EMBL: AF000003; BAA29946.1;
DR InterPro: IPR003032;
DR InterPro: IPR003031;
DR Pfam: PF00206; Lyase_1; 1.
DR PRINTS; PR00145; DCRYSTALLIN.
DR PRINTS; PR00149; FUMARATE_LYASE.
DR PROSITE: PS00163; FUMARATE_LYASES; 1.
KW Purine biosynthesis; Lyase.
FT ACT_SITE 76 76 ACID (BY SIMILARITY).
FT ACT_SITE 149 149 BASE (BY SIMILARITY).
SQ SEQUENCE 450 AA; 51644 MW; A700A652ADB822BC CRC64;

Query Match 51.8%; Score 43; DB 1; Length 450;
Best Local Similarity 64.3%; Pred. No. 16;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 LQGWGGAAGTAA 16
   |:||| | |
Db 184 LVGKMGAVGTAAS 197

RESULT 6
SPDL_NEPCL STANDARD; PRT; 747 AA.
AC PI9837;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
```





CC -1- DOMAIN: COMPOSED OF AN N- AND A C-TERMINAL DOMAIN. THE N-TERMINAL  
CC DOMAIN CARRIES THE TETRAHYDROFOLATE (THF)-BINDING SITE AND THE C-  
CC TERMINAL DOMAIN IS PRESUMABLY INVOLVED IN POSITIONING THE MET-TRNA  
CC SUBSTRATE FOR THE FORMYLATION REACTION.  
CC -1- SIMILARITY: BELONGS TO THE FMT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X63666; CAA45207.1; -  
DR EMBL: X00767; CAA25339.1; ALT\_SEQ.  
DR EMBL: X77091; CAA34368.1; -  
DR EMBL: U18997; AAA58085.1; -  
DR EMBL: Y10307; CAA71358.1; -  
DR EMBL: AE000407; AAC76313.1; -  
DR PIR: S23108; S23108.  
DR PDB: 1FMT; 28-JAN-98.  
DR SWISS-2DPAGE; P23882; COLI.  
DR ECO2DBASE; F033.6; 6TH EDITION.  
DR EcoGene; EGI1268; fnt.  
DR InterPro; IPR001555; -  
DR InterPro; IPR002376; -  
DR Pfam; PF00551; formyl\_transf; 1.  
DR PROSITE; PS00373; GART; 1.  
KW Transferase; Methyltransferase; Protein biosynthesis; 3D-structure.  
FT INIT\_MET 0  
FT DOMAIN 1 188 N-TERMINAL DOMAIN.  
FT DOMAIN 209 314 C-TERMINAL DOMAIN.  
FT BINDING 112 115 TETRAHYDROFOLATE (THF).  
SQ SEQUENCE 314 AA; 34037 MW; 83D8AE62B616A1C4 CRC64;  
  
Query Match 49.4%; Score 41; DB 1; Length 314;  
Best Local Similarity 72.7%; Pred. No. 23;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 GSLQGWGAA 11  
DB 111 GLLPRWGAA 121  
  
RESULT 8  
TRBL\_AERPE STANDARD; PRT; 427 AA.  
AC Q9Y8T5;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE TRYPTOPHAN SYNTHASE BETA CHAIN 1 (EC 4.2.1.20).  
GN TRPB1 OR APE2548.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Desulfurococcaceae;  
OC Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kavaravasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,  
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic  
RT crenarchaeon, Aeropyrum pernix K1.";  
RL DNA Res. 6:83-101(1999).  
CC -1- FUNCTION: THE BETA SUBUNIT IS RESPONSIBLE FOR THE SYNTHESIS OF L-

CC TRYPTOPHAN FROM INDOLE AND L-SERINE.  
CC -1- CATALYTIC ACTIVITY: L-SERINE + 1-(INDOL-3-YL)GLYCEROL 3-PHOSPHATE  
CC = L-TRYPTOPHAN + GLYCERALDEHYDE 3-PHOSPHATE + H(2)O.  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
CC -1- PATHWAY: LAST (FIFTH) STEP IN BIOSYNTHESIS OF TRYPTOPHAN.  
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE TRPB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AP000064; BAA81565.1; -  
DR InterPro; IPR001926; -  
DR Pfam; PF00291; S\_T\_dehydratase; 1.  
DR PROSITE; PS00168; TRP\_SYNTHASE\_BETA; 1.  
KW Tryptophan biosynthesis; Pyridoxal phosphate; Lyase.  
FT BINDING 107 107 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
SQ SEQUENCE 427 AA; 45992 MW; DC8E165840B65750 CRC64;  
  
Query Match 49.4%; Score 41; DB 1; Length 427;  
Best Local Similarity 75.0%; Pred. No. 30;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 5 GQWRGAAGTAAA 16  
DB 133 GQWGLAASPTAA 144  
  
RESULT 9  
YJIW\_ECOLI STANDARD; PRT; 132 AA.  
ID YJIW\_ECOLI  
AC P39394;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 14.6 KDA PROTEIN IN MCRB-HSDS INTERGENIC REGION (F132).  
GN YJIW.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=95334362; PubMed=7610040;  
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
RA Blattner F.R.;  
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
RT region from 92.8 through 100 minutes.";  
RL Nucleic Acids Res. 23:2105-2119(1995).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U14003; AAA97244.1; -  
DR EMBL: AE000505; AAC77303.1; -  
DR EcoGene; EGI2584; YJIW.  
KW Hypothetical protein.  
SQ SEQUENCE 132 AA; 14576 MW; 3E53097CD17B0C62 CRC64;

Query Match 48.28; Score 40; DB 1; Length 132;  
Best Local Similarity 61.54; Pred. No. 15;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 SLOGOWRGAAGTA 14  
:||||| |||  
Db 60 TLKGWLEAAGTA 72

RESULT 10  
COX1\_ORNAN STANDARD; PRT; 513 AA.  
AC Q36452;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
GN MTCO1 OR COI OR COXI.  
OS Ornithorhynchus anatinus (Duckbill platypus).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.  
OX NCBI\_TaxID=9258;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart, and Liver;  
RX MEDLINE=97077300; PubMed=8919867;  
RA Janke A., Gemmell N., Feldmaier-Fuchs G., von Haeseler A.,  
RA Paabo S.;  
RT "The mitochondrial genome of a monotreme -- the platypus  
RT (Ornithorhynchus anatinus).";  
RL J. Mol. Evol. 42:153-159(1996).  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1 -  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
CC AND COPPER B.  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) - 2 H(2)O +  
CC 4 FERRICYTOCHROME C.  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X83427; CAA58457.1; -  
CC HSPF; P98002; IAR1.  
CC InterPro; IPR000883; -  
CC InterPro; IPR002428; -  
CC Pfam; PF00115; COX1; 1.  
CC PRINTS; PR01165; CYCOXIDASE1.  
CC PROSITE; PS00077; COX1; 1.  
CC Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
KW Respiratory chain; Inner membrane.  
FT METAL 61 61 IRON (HEME A) (PROBABLE).  
FT METAL 240 240 COPPER B (PROBABLE).  
FT METAL 244 244 COPPER B (PROBABLE).  
FT METAL 290 290 COPPER B (PROBABLE).  
FT METAL 291 291 COPPER B (PROBABLE).  
FT METAL 376 376 IRON (HEME A3) (PROBABLE).  
FT METAL 378 378 IRON (HEME A) (PROBABLE).  
SQ SEQUENCE 513 AA; 56742 MW; BC8758C005810841 CRC64;

Query Match 48.28; Score 40; DB 1; Length 513;  
Best Local Similarity 66.74; Pred. No. 50;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 LOGOWRGAAGTA 14  
:||||| |||  
Db 21 LFGAWAGTA 32

RESULT 11  
NIEK\_BRASP STANDARD; PRT; 513 AA.  
AC P06122;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE NITROGENASE MOLYBDENUM-IRON PROTEIN BETA CHAIN (EC 1.18.6.1)  
DE (NITROGENASE COMPONENT I) (DINITROGENASE).  
GN NIFK.  
OS Bradyrhizobium sp. (strain ANU 289).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Bradyrhizobium group; Bradyrhizobium.  
OX NCBI\_TaxID=377;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85062817; PubMed=6095197;  
RA Weinman J.J., Fellows F.F., Gresshoff P.M., Shine J., Scott K.F.;  
RT "Structural analysis of the genes encoding the molybdenum-iron  
RT protein of nitrogenase in the Parasponia rhizobium strain ANU289.";  
RL Nucleic Acids Res. 12:8329-8344(1984).  
CC -1- FUNCTION: THE KEY ENZYMAIC REACTIONS IN NITROGEN FIXATION ARE  
CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE  
CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.  
CC -1- CATALYTIC ACTIVITY: 8 REDUCED FERREDOXIN + 8 H(+) + N(2) + 16 ATP  
CC - 8 OXIDIZED FERREDOXIN + 2 NH(3) + 16 ADP + 16 ORTHOPHOSPHATE.  
CC -1- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS THAT BINDS  
CC 30-32 FE, 2 MO, AND INORGANIC SULFUR.  
CC -1- SIMILARITY: BELONGS TO THE NIFD/NIEK/NIFE/NIFEN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X01139; CAA25597.1; -  
CC PIR; B23874; B23874.  
CC HSPF; P07329; IN2C.  
CC InterPro; IPR000318; -  
CC InterPro; IPR000510; -  
CC Pfam; PF00148; Oxidored\_nitro; 1.  
CC PROSITE; PS00090; NITROGENASE\_1\_2; 1.  
CC PROSITE; PS00699; NITROGENASE\_1\_1; 1.  
KW Oxidoreductase; Nitrogen fixation; Molybdenum; Iron-sulfur.  
SQ SEQUENCE 513 AA; 56538 MW; 741373586CDD8A36 CRC64;

Query Match 48.28; Score 40; DB 1; Length 513;  
Best Local Similarity 57.14; Pred. No. 50;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 GSLOGOWRGAAGTA 14  
:||||| |||  
Db 199 GILEHFWNGRAGTA 212

RESULT 12  
COX1\_HALGR STANDARD; PRT; 514 AA.  
ID COX1\_HALGR  
AC P38595;  
DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (rel. 30, Last sequence update)  
DT 01-NOV-1997 (rel. 35, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).  
GN MTCOI OR COL.  
OS Halichoerus grypus (Gray seal).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Pinnipedia; Phocidae; Halichoerus.  
OX NCBI\_TaxID=9711;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94141933; PubMed=8308902;  
RA Arnason U., Gullberg A., Johnsson E., Ledje C.;  
RT "The nucleotide sequence of the mitochondrial DNA molecule of the  
RT grey seal, Halichoerus grypus, and a comparison with mitochondrial  
RT sequences of other true seals."  
RL J. Mol. Evol. 37:323-330(1993).  
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3  
CC AND COPPER B.  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) - 2 H(2)O +  
CC 4 FERRICYTOCHROME C.  
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, MITOCHONDRIAL  
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X72004; CAA50879.1; -  
DR PIR: S41837; S41837.  
DR HSSP: P00396; LOCC.  
DR InterPro: IPR000883; -  
DR InterPro: IPR002428; -  
DR Pfam: PF00115; COX1; 1.  
DR PRINTS: PR01165; CYCOXIDASE1.  
DR PROSITE: PS00077; COX1; 1.  
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
KW Respiratory chain; Inner membrane.  
FT METAL 61 61 IRON (HEME A) (PROBABLE).  
FT METAL 240 240 COPPER B (PROBABLE).  
FT METAL 244 244 COPPER B (PROBABLE).  
FT METAL 290 290 COPPER B (PROBABLE).  
FT METAL 291 291 COPPER B (PROBABLE).  
FT METAL 376 376 IRON (HEME A3) (PROBABLE).  
FT METAL 378 378 IRON (HEME A3) (PROBABLE).  
SQ SEQUENCE 514 AA; 56925 MW; 2A40D55A0140F233 CRC64;  
  
Query Match 48.2%; Score 40; DB 1; Length 514;  
Best Local Similarity 66.7%; Pred. No. 50;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 3 LOGQWRGAAGTA 14  
| | | | |  
Db 21 LFQWAGMAGTA 32  
  
RESULT 13  
NXB2\_CERLA  
ID NXB2\_CERLA STANDARD; PRT; 55 AA.  
AC P01526;  
DT 21-JUL-1986 (Rel. 01, Created)  
DE Cerebratulus lacteus (Milky ribbon worm).  
OC Eukaryota; Metazoa; Nemertea; Anopla; Heteronemertea; Cerebratulus.  
OX NCBI\_TaxID=6221;  
RN [1]  
RP SEQUENCE, AND REVISIONS.  
RX MEDLINE=81264350; PubMed=7263698;  
RA Blumenthal K.M., Keim P.S., Heinrichson R.L., Kem W.R.;  
RT "Structure and action of heteronemertine polypeptide toxins. Amino  
RT acid sequence of Cerebratulus lacteus toxin B-II and revised  
RT structure of toxin B-IV.";  
RL J. Biol. Chem. 256:9063-9067(1981).  
RN [2]  
RP SEQUENCE.  
RX MEDLINE=77006160; PubMed=972152;  
RA Blumenthal K.M., Kem W.R.;  
RT "Structure and action of heteronemertine polypeptide toxins. Primary  
RT structure of Cerebratulus lacteus toxin B-IV.";  
RL J. Biol. Chem. 251:6025-6029(1976).  
RN [3]  
RP MUTAGENESIS.  
RX MEDLINE=91302300; PubMed=2071577;  
RA Howell M.L., Blumenthal K.M.;  
RT "Mutagenesis of Cerebratulus lacteus neurotoxin B-IV identifies NH2-  
RT terminal sequences important for biological activity.";  
RL J. Biol. Chem. 266:12884-12888(1991).  
RN [4]  
RP STRUCTURE BY NMR.  
RX MEDLINE=93076792; PubMed=1332861;  
RA Hansen P.E., Kem W.R., Bieber A.L., Norton R.S.;  
RT "1H-NMR study of neurotoxin B-IV from the marine worm Cerebratulus

DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 01-AUG-1991 (Rel. 19, Last annotation update)  
DE NEUROTOXIN B-II.  
OS Cerebratulus lacteus (Milky ribbon worm).  
OC Eukaryota; Metazoa; Nemertea; Anopla; Heteronemertea; Cerebratulus.  
OX NCBI\_TaxID=6221;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=81264350; PubMed=7263698;  
RA Blumenthal K.M., Keim P.S., Heinrichson R.L., Kem W.R.;  
RT "Structure and action of heteronemertine polypeptide toxins. Amino  
RT acid sequence of Cerebratulus lacteus toxin B-II and revised  
RT structure of toxin B-IV.";  
RL J. Biol. Chem. 256:9063-9067(1981).  
RN [2]  
RP SEQUENCE.  
RX MEDLINE=77006160; PubMed=972152;  
RA Blumenthal K.M., Kem W.R.;  
RT "Structure and action of heteronemertine polypeptide toxins. Primary  
RT structure of Cerebratulus lacteus toxin B-IV.";  
RL J. Biol. Chem. 251:6025-6029(1976).  
RN [3]  
RP MUTAGENESIS.  
RX MEDLINE=91302300; PubMed=2071577;  
RA Howell M.L., Blumenthal K.M.;  
RT "Mutagenesis of Cerebratulus lacteus neurotoxin B-IV identifies NH2-  
RT terminal sequences important for biological activity.";  
RL J. Biol. Chem. 266:12884-12888(1991).  
RN [4]  
RP STRUCTURE BY NMR.  
RX MEDLINE=93076792; PubMed=1332861;  
RA Hansen P.E., Kem W.R., Bieber A.L., Norton R.S.;  
RT "1H-NMR study of neurotoxin B-IV from the marine worm Cerebratulus

Query Match 47.0%; Score 39; DB 1; Length 55;  
Best Local Similarity 53.8%; Pred. No. 9.5;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 4 QGQWRGAAGTAA 16  
| | | | |  
Db 27 QGKWKAGKRGKCA 39  
  
RESULT 14  
NXB4\_CERLA  
ID NXB4\_CERLA STANDARD; PRT; 55 AA.  
AC P01525;  
DT 21-JUL-1986 (Rel. 01, Created)  
DE Cerebratulus lacteus (Milky ribbon worm).  
OC Eukaryota; Metazoa; Nemertea; Anopla; Heteronemertea; Cerebratulus.  
OX NCBI\_TaxID=6221;  
RN [1]  
RP SEQUENCE, AND REVISIONS.  
RX MEDLINE=81264350; PubMed=7263698;  
RA Blumenthal K.M., Keim P.S., Heinrichson R.L., Kem W.R.;  
RT "Structure and action of heteronemertine polypeptide toxins. Amino  
RT acid sequence of Cerebratulus lacteus toxin B-II and revised  
RT structure of toxin B-IV.";  
RL J. Biol. Chem. 256:9063-9067(1981).  
RN [2]  
RP SEQUENCE.  
RX MEDLINE=77006160; PubMed=972152;  
RA Blumenthal K.M., Kem W.R.;  
RT "Structure and action of heteronemertine polypeptide toxins. Primary  
RT structure of Cerebratulus lacteus toxin B-IV.";  
RL J. Biol. Chem. 251:6025-6029(1976).  
RN [3]  
RP MUTAGENESIS.  
RX MEDLINE=91302300; PubMed=2071577;  
RA Howell M.L., Blumenthal K.M.;  
RT "Mutagenesis of Cerebratulus lacteus neurotoxin B-IV identifies NH2-  
RT terminal sequences important for biological activity.";  
RL J. Biol. Chem. 266:12884-12888(1991).  
RN [4]  
RP STRUCTURE BY NMR.  
RX MEDLINE=93076792; PubMed=1332861;  
RA Hansen P.E., Kem W.R., Bieber A.L., Norton R.S.;  
RT "1H-NMR study of neurotoxin B-IV from the marine worm Cerebratulus

RT lacteus. Solution properties, sequence-specific resonance  
RL assignments, secondary structure and global fold.";  
RN Eur. J. Biochem. 210:231-240(1992).  
RP STRUCTURE BY NMR.  
RX MEDLINE-97324210; PubMed-9180379;  
RA Barnham K.J., Dyke T.R., Kem W.R., Norton R.S.;  
RT "Structure of neurotoxin B-IV from the marine worm Cerebratulus  
lacteus: a helical hairpin cross-linked by disulphide bonding.";  
RL J. Mol. Biol. 268:886-902(1997).  
CC -1- FUNCTION: ONLY ACTS ON SOME CRUSTACEAN. INCREASE THE EXCITABILITY  
OF NERVES PROBABLY BY AFFECTING THE INACTIVATION OF THE VOLTAGE-  
GATED SODIUM CHANNEL.  
CC -1- SIMILARITY: TO NEUROTOXIN B-II.  
DR PIR: A01788; NTHNB4.  
DR PDB: 1VIB; 15-MAY-97.  
KW Toxin; Hydroxylation; 3D-structure.  
FT MOD\_RES 10 10 HYDROXYLATION.  
FT DISULFID 12 48  
FT DISULFID 16 52  
FT DISULFID 23 41  
FT DISULFID 26 37  
SQ SEQUENCE 55 AA; 6107 MW; BB76B72E48DB050D CRC64;

Query Match 47.0%; Score 39; DB 1; Length 55;  
Best Local Similarity 53.8%; Pred. No. 9.5;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 4 QGWRGAAGTAA 16  
||| | | |  
Db 27 QGKWAGKRGKCA 39

RESULT 15  
COXI\_CORJA  
ID COXI\_COTJA STANDARD; PRT; 156 AA.  
AC P24984;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENTS).  
GN MTCOI OR COI.  
OS Coturnix coturnix japonica (Japanese quail).  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Coturnix.  
OX NCBI\_TaxID=93934;  
RN [1]  
RP SEQUENCE OF 1-102 FROM N.A.  
RP TISSUE=Liver;  
RX MEDLINE-91178819; PubMed-1706782;  
RA Desjardins P., Morais R.;  
RT "Nucleotide sequence and evolution of coding and noncoding regions of  
a quail mitochondrial genome.";  
RL J. Mol. Evol. 32:153-161(1991).  
RN [2]  
RP SEQUENCE OF 103-156 FROM N.A.  
RP TISSUE=Liver;  
RA Ramirez V., Morais R.;  
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-  
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE  
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN  
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2  
AND HEME A OF SUBUNIT 1 TO THE COPPER A CENTER FORMED BY HEME A3  
AND COPPER B.  
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +  
4 FERRICYTOCHROME C.  
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL

CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X57246; CAA40524.1; -  
DR EMBL; U36794; AAA76729.1; -  
DR PIR; S25424; S25424.  
DR HSP; P00396; LOCC.  
DR InterPro: IPR000883; -  
DR Pfam: PF00115; COXI; 1.  
DR PROSITE; PS00077; COXI; PARTIAL.  
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;  
KW Respiratory chain; Inner membrane.  
FT METAL 62 62 IRON (HEME A) (PROBABLE).  
FT NON\_CONS 102 103  
SQ SEQUENCE 156 AA; 17502 MW; C18A3282896190AA CRC64;

Query Match 47.0%; Score 39; DB 1; Length 156;  
Best Local Similarity 70.0%; Pred. No. 24;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 GQWRGAAGTA 14  
| | | | |  
Db 24 GTWAGMAGTA 33

Search completed: July 5, 2001, 11:51:44  
Job time: 503 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:50:50 ; Search time 123.78 Seconds  
(without alignments)  
17.102 Million cell updates/sec

Title: US-09-462-480-10

Perfect score: 83

Sequence: 1 GSLOGQWRGAAGTAA 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_16:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	79	95.2	100	2	069739		069739 mycobacteri
2	48	57.8	103	2	005440		005440 mycobacteri
3	47	56.6	128	2	086441		086441 pseudomonas
4	47	56.6	592	6	09XSR3		09XSR3 canis famli
5	45	54.2	291	2	09EWV9		09EWV9 streptomyce
6	44	53.0	485	10	09FTS8		09FTS8 oryza sativ
7	44	53.0	656	2	P72045		P72045 mycobacteri
8	44	53.0	1186	2	09ZB02		09ZB02 streptomyce
9	43	51.8	124	2	086643		086643 streptomyce
10	43	51.8	203	11	09ZLS6		09ZLS6 rattus norv
11	43	51.8	314	2	09L0Q5		09L0Q5 streptomyce
12	43	51.8	331	2	09X7H6		09X7H6 paracoccus
13	43	51.8	391	11	088676		088676 mus musculu
14	43	51.8	542	2	086368		086368 mycobacteri
15	42	50.6	179	10	09M0L8		09M0L8 arabidopsis
16	42	50.6	277	10	049678		049678 arabidopsis
17	42	50.6	366	2	Q44163		Q44163 actinobacil
18	42	50.6	367	2	Q44164		Q44164 actinobacil
19	42	50.6	367	2	Q9ZAP9		Q9ZAP9 actinobacil

20 20 42 50.6 368 2 09RIH8  
21 42 50.6 368 2 09RCG3  
22 42 50.6 442 4 095932  
23 42 50.6 522 10 09LIW0  
24 42 50.6 533 10 004400  
25 42 50.6 544 5 046171  
26 42 50.6 617 5 046172  
27 42 50.6 650 10 09XIX5  
28 41 49.4 69 6 09N0B6  
29 41 49.4 181 10 09XEX9  
30 41 49.4 205 13 042394  
31 41 49.4 230 8 079121  
32 41 49.4 249 8 09XM88  
33 41 49.4 249 8 09XM87  
34 41 49.4 284 2 066177  
35 41 49.4 288 2 09Z483  
36 41 49.4 290 2 09F373  
37 41 49.4 294 2 007242  
38 41 49.4 361 14 098307  
39 41 49.4 369 10 09XE89  
40 41 49.4 381 2 09I3E5  
41 41 49.4 403 10 09SNH4  
42 41 49.4 513 8 099040  
43 41 49.4 528 5 060978  
44 41 49.4 605 11 088882  
45 41 49.4 611 11 088881

#### ALIGNMENTS

RESULT 1  
ID 069739 PRELIMINARY; PRT; 100 AA.  
AC 069739;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE HYPOPHETICAL 10.8 KDA PROTEIN.  
GN LHP OR RV3874 OR MTV027.09.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RA Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;  
RT "Promoter analysis of the M. tuberculosis orf1C gene encoding the  
RT early secreted antigenic target 6 kDa (ESAT-6).";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF002120; CAAL1966.1; -;  
DR EMBL; AF004671; AAC83445.1; -;  
DR TubercuList; RV3874; -;  
KW Hypothetical protein.  
SQ SEQUENCE 100 AA; 10794 MW; 285F4FC96F55D194 CRC64;

Query Match 95.2%; Score 79; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 6.1e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAA 15  
| | | | | | | | | | | | | | | | |  
DB 37 GSLOGQWRGAAGTAA 51

RESULT 2  
ID O05440 PRELIMINARY; PRT; 103 AA.  
AC O05440;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE HYPOTHETICAL 10.5 KDA PROTEIN.  
GN RV3905C OR MTCY15F10.06.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
DR EMBL; Z94121; CAB08096.1; -  
DR TubercuList; RV3905C; -  
KW Hypothetical protein.  
SQ SEQUENCE 103 AA; 10460 MW; 3994E272A7BDF02 CRC64;

Query Match 57.8%; Score 48; DB 2; Length 103;  
Best Local Similarity 57.1%; Pred. No. 3;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAA 14  
| | | | | | | | | | | | | | | | |  
DB 40 GMLGGWRGASGSA 53

RESULT 3  
ID O86441 PRELIMINARY; PRT; 128 AA.  
AC O86441;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE XCP2 PROTEIN.  
GN XCP2.  
OS Pseudomonas putida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WCS358;  
RA de Groot A., Gerritse G., Lazdunski A., Filloux A.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X81085; CAA56987.1; -  
SQ SEQUENCE 128 AA; 14154 MW; A32A79E69EAB6F18 CRC64;

Query Match 56.6%; Score 47; DB 2; Length 128;  
Best Local Similarity 81.8%; Pred. No. 5.4;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GQWRGAAGTAA 15  
| | | | | | | | | | | | | | | | |  
DB 38 GQWRGLAQTA 48

RESULT 4  
ID Q9XSR3 PRELIMINARY; PRT; 592 AA.  
AC Q9XSR3;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE HYPOTHETICAL 63.5 KDA PROTEIN.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=THYROID;  
RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,  
RA Christophe D.;  
RT "Functional cloning of nuclear proteins and nuclear targeting  
RT sequences.";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ388555; CAB46854.1; -  
DR HSP; P09651; IHAL.  
DR InterPro; IPR000504; -  
DR Pfam; PF00076; rim; 3.  
DR PROSITE; PS00030; RNP\_1; UNKNOWN\_2.  
DR SMART; SM00360; RM; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 592 AA; 63523 MW; 5952B93616CBE4E5 CRC64;

Query Match 56.6%; Score 47; DB 6; Length 592;  
Best Local Similarity 60.0%; Pred. No. 25;  
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 GSLOGQWRGAAGTAA 15  
| | | | | | | | | | | | | | | | |  
DB 379 GSIRGRGRGAAGSRA 393

RESULT 5  
ID Q9EWV9 PRELIMINARY; PRT; 291 AA.  
AC Q9EWV9;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PUTATIVE LIGASE.  
GN 25CK31.14C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]



```
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
KW EMBL; AL512667; CAC21623.1; -.
KW Ligase.
SQ SEQUENCE 291 AA; 30545 MW; 7C986648485F87C6F CRC64;

Query Match 54.2%; Score 45; DB 2; Length 291;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 GQWRGAAGTAA 16
   :||||| |
DB 216 GWRGAGDPAA 227

RESULT 6
Q9FTS8 PRELIMINARY; PRT; 485 AA.
AC Q9FTS8;
DT 01-WAR-2001 (TREMELREL. 16, Created)
DT 01-WAR-2001 (TREMELREL. 16, Last sequence update)
DT 01-WAR-2001 (TREMELREL. 16, Last annotation update)
DE P0409B08.8 PROTEIN.
GN P0409B08.8.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0409B08.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002860; BAB18282.1; -.
SQ SEQUENCE 485 AA; 50477 MW; 091B5D3AF14DF5FA CRC64;

Query Match 53.0%; Score 44; DB 10; Length 485;
Best Local Similarity 53.3%; Pred. No. 59;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 GSLOGOWRGAAGTAA 15
   | | : ||| | |
DB 284 GGLRAEWRGAWTA 298

RESULT 7
P72045 PRELIMINARY; PRT; 566 AA.
ID P72045;
AC P72045;
DT 01-JAN-1998 (TREMELREL. 05, Created)
DT 01-JAN-1998 (TREMELREL. 05, Last sequence update)
DT 01-JUN-2000 (TREMELREL. 14, Last annotation update)
DE HYPOTHETICAL 71.8 KDA PROTEIN CY13D12.13.
GN RV3779 OR WTCY13D12.13.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A.;
```

```
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
DR EMBL; Z80343; CAB02458.1; -.
DR Tuberculiist; RV3779; -.
DR InterPro; IPR001084; -.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 97 117 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 292 312 POTENTIAL.
FT TRANSMEM 333 353 POTENTIAL.
FT TRANSMEM 404 424 POTENTIAL.
FT TRANSMEM 428 448 POTENTIAL.
FT TRANSMEM 460 480 POTENTIAL.
FT TRANSMEM 493 513 POTENTIAL.
SQ SEQUENCE 666 AA; 71794 MW; 071B932A1BDC9BB0 CRC64;

Query Match 53.0%; Score 44; DB 2; Length 666;
Best Local Similarity 76.9%; Pred. No. 81;
Matches 10; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 6 QWR--GAAGTAA 16
   :| | | | | | | |
DB 231 EWRAAGAAGTAA 243

RESULT 8
Q9ZBQ2 PRELIMINARY; PRT; 1186 AA.
ID Q9ZBQ2;
AC Q9ZBQ2;
DT 01-MAY-1999 (TREMELREL. 10, Created)
DT 01-MAY-1999 (TREMELREL. 10, Last sequence update)
DT 01-WAR-2001 (TREMELREL. 16, Last annotation update)
DE PUTATIVE CHROMOSOME ASSOCIATED PROTEIN.
GN SC7A1.21.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL034447; CAA22420.1; -.
DR InterPro; IPR001066; -.
DR InterPro; IPR001687; -.
DR InterPro; IPR003395; -.
DR InterPro; IPR003405; -.
DR InterPro; IPR003439; -.
DR Pfam; PF02463; SMC_N; 1.
DR Pfam; PF02483; SMC_C; 1.
```

DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
SQ SEQUENCE 1186 AA; 128722 MW; CB11027815373E99 CRC64;

Query Match 53.0%; Score 44; DB 2; Length 1186;  
Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTA 14  
| | | | |  
Db 718 GRLAQQRGAAGEA 731

## RESULT 9

ID O86643 PRELIMINARY; PRT; 124 AA.  
AC O86643;  
DT 01-NOV-1998 (TRENBLrel. 08, Created)  
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)  
DE HYPOTHEICAL 13.9 KDA PROTEIN.  
GN SC3C3.10.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
FN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Murphy L., Harris D.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL MOL. Microbiol. 21:77-96(1996).  
DR EMBL; AL031231; CAA20259.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 124 AA; 13867 MW; F5E17CF4C5949FFA CRC64;

Query Match 51.8%; Score 43; DB 2; Length 124;  
Best Local Similarity 70.0%; Pred. No. 21;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 LOGQWRGAAG 12  
| | | | |  
Db 52 IQGGWQGAAG 61

## RESULT 10

ID Q921S6 PRELIMINARY; PRT; 203 AA.  
AC Q921S6;  
DT 01-MAY-1999 (TRENBLrel. 10, Created)  
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)  
DE MELANOCYTE-SPECIFIC PROTEIN 1.  
GN MSG1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN (1)

RP SEQUENCE FROM N.A.  
RA Plisov S.Y., Ivanov S.V., Lerman M., Perantoni A.O.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF104399; AAC98389.1; -  
SQ SEQUENCE 203 AA; 20695 MW; 9DDA634E80303B32 CRC64;

Query Match 51.8%; Score 43; DB 11; Length 203;  
Best Local Similarity 64.3%; Pred. No. 34;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 LOGQWRGAAGTAA 16  
| | | | |  
Db 102 LNSOYQGAATATA 115

## RESULT 11

ID Q9L0Q5 PRELIMINARY; PRT; 314 AA.  
AC Q9L0Q5;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
DE PUTATIVE TRNA/RNA METHYLTRANSFERASE.  
GN SCD8A.09.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL MOL. Microbiol. 21:77-96(1996).  
DR EMBL; AL160331; CAB77330.1; -  
DR InterPro; IPR001537; -  
DR Pfam; PF00588; SpoU\_methylase; 1.  
KW Transferase; Methyltransferase.  
SQ SEQUENCE 314 AA; 32967 MW; 609F43543485C5DD CRC64;

Query Match 51.8%; Score 43; DB 2; Length 314;  
Best Local Similarity 53.3%; Pred. No. 53;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GSLOGQWRGAAGTAA 15  
| | | | |  
Db 203 GMTAGAKTSAGTAA 217

## RESULT 12

ID Q9X7H6 PRELIMINARY; PRT; 331 AA.  
AC Q9X7H6;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
DE GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12).  
GN GAP.

OS Paracoccus denitrificans.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Paracoccus.  
OX NCBI\_TaxID=266;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99261655; PubMed=10331270;  
RA Fligge R.M., Schubert M., Brinkmann H., Cerff R.;  
RT "Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria  
and eukaryotes: evidence for intra- and inter-kingdom gene transfer.";  
RL Mol. Biol. Evol. 16:429-440(1999).  
CC -1- CATALYTIC ACTIVITY: ICCA D-GLYCERALDEHYDE 3-PHOSPHATE +  
ORTHOPHOSPHATE + NAD(+) -> 1,3-DIPHOSPHATEGLYCERATE + NADH (BY  
SIMILARITY).  
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS (BY  
SIMILARITY).  
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE  
DEHYDROGENASE FAMILY.  
DR EMBL: AJ012158; CAB41844.1; -.  
DR HSSP: P17721; 1HDG.  
DR InterPro: IPR000173; -.  
DR Pfam: PF00044; gpdh; 1.  
DR PRINTS: PR00078; G3PDHGRNASE.  
DR PROSITE: PS00071; GAPDH; UNKNOWN\_1.  
KW Glycolysis; NAD; Oxidoreductase.  
SQ SEQUENCE 331 AA; 35061 MW; B93325B171338C59 CRC64;  
  
Query Match 51.8%; Score 43; DB 2; Length 331;  
Best Local Similarity 61.5%; Pred. No. 56;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 SLOGQWRGAGTA 14  
Db 54 SVFGPWRGQVGT 66  
|:|||||  
  
RESULT 13  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=SPLEEN;  
RA Pei D.;  
RT "CAMP, a matrix metalloproteinase without a classic cysteine switch.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX  
DOMAIN.  
DR EMBL: AF085742; AAC34886.1; -.  
DR HSSP: P03956; 1CGL.  
DR MEROPS: M10.022; -.  
DR MGD: MGI:1347361; Mmp23.  
DR InterPro: IPR000130; -.  
DR InterPro: IPR000566; -.  
DR InterPro: IPR001818; -.  
DR InterPro: IPR003006; -.  
DR InterPro: IPR003599; -.  
DR Pfam: PF00047; ig; 1.  
DR Pfam: PF00413; Peptidase\_M10; 1.  
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN\_1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; UNKNOWN\_1.

DR SMART: SM00409; IG; 1.  
SQ SEQUENCE 391 AA; 44451 MW; 8C9675020F02F632 CRC64;  
  
Query Match 51.8%; Score 43; DB 11; Length 391;  
Best Local Similarity 63.6%; Pred. No. 67;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GSLOGQWRGAA 11  
Db 13 GAVQGRWLGA 23  
|:|||||  
  
RESULT 14  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
RA Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
RT Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RA Parkhill J.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL021943; CAAL7469.1; -.  
DR Tuberculist; RV0672; -.  
DR InterPro: IPR001552; -.  
DR Pfam: PF00441; Acyl-CoA\_dh; 1.  
DR PROSITE: PS00073; ACYL\_COA\_DH\_2; UNKNOWN\_1.  
SQ SEQUENCE 542 AA; 58534 MW; 1AD435206E5E2C33 CRC64;  
  
Query Match 51.8%; Score 43; DB 2; Length 542;  
Best Local Similarity 64.3%; Pred. No. 93;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
  
QY 3 LGGQWRGAGTA 16  
Db 514 LGGQWRGAGT 527  
||||||  
  
RESULT 15  
ID Q9M0L8  
AC Q9M0L8; PRELIMINARY; PRT; 179 AA.  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 17.7 KDA PROTEIN.  
GN AT4G19200.  
OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Benes V., Rechmann S., Borkova D., Ansoerge W., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL161550; CAB78922.1; -  
DR InterPro; IPR000216; -  
DR PRINTS; PR00239; RHODOPSNTAIL.  
KW Hypothetical protein.  
SQ SEQUENCE 179 AA; 17745 MW; 1DAC3F3D3C1A604D CRC64;

Query Match 50.6%; Score 42; DB 10; Length 179;  
Best Local Similarity 62.5%; Pred. No. 43;  
Matches 10; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 GSLOGQWRGAAGTAA 16  
| | | | |  
Db 91 GGLGMIAGAAGAAA 106

Search completed: July 5, 2001, 11:50:51  
Job time: 474 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:45:43 ; Search time 130.35 seconds  
(without alignments)  
7.441 Million cell updates/sec

Title: US-09-462-480-9

Perfect score: 78

Sequence: 1 GDLKTQIDQVESTAGS 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_0601:\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	100.0	16	AAV03709	M. tuberculosis LH
2	78	100.0	49	AAV03706	M. tuberculosis LH
3	78	100.0	80	AAW32454	Mycobacterium tube
4	78	100.0	80	AAW32386	Mycobacterium tube
5	78	100.0	80	AAW81707	M. tuberculosis im
6	78	100.0	80	AAW64340	Mycobacterium tube
7	78	100.0	80	AAV39137	M. tuberculosis an
8	78	100.0	80	AAV38994	M. tuberculosis re
9	78	100.0	95	AAW32444	Mycobacterium tube
10	78	100.0	95	AAW32376	Mycobacterium tube
11	78	100.0	95	AAW81747	M. tuberculosis im

12	78	100.0	95	19	AAW64321	Mycobacterium tube
13	78	100.0	95	20	AAV32097	Mycobacterium tube
14	78	100.0	95	20	AAV39118	M. tuberculosis an
15	78	100.0	95	20	AAV38981	M. tuberculosis re
16	78	100.0	100	19	AAW81706	Mycobacterium im
17	78	100.0	100	19	AAW64339	Mycobacterium tube
18	78	100.0	100	20	AAV39136	M. tuberculosis an
19	78	100.0	100	20	AAV38993	M. tuberculosis re
20	78	100.0	100	20	AAV03705	M. tuberculosis LH
21	78	100.0	100	22	AAW35218	M. tuberculosis RV3
22	78	100.0	100	22	AAW19845	Mycobacterium tube
23	78	100.0	802	19	AAW81746	M. tuberculosis fu
24	78	100.0	802	19	AAW64379	Mycobacterium anti
25	78	100.0	802	20	AAV32063	Mycobacterium tube
26	78	100.0	802	20	AAV39224	M. tuberculosis fu
27	78	100.0	802	20	AAV39176	M. tuberculosis fu
28	78	100.0	802	20	AAV39081	M. tuberculosis fus
29	78	100.0	802	20	AAV39033	M. tuberculosis
30	50	64.1	20	20	AAV03708	M. tuberculosis
31	46	59.0	778	21	AAW51229	Arabisopsis thalia
32	46	59.0	806	21	AAW51228	Arabisopsis thalia
33	46	59.0	927	21	AAW51227	Arabisopsis thalia
34	46	59.0	1053	21	AAW50491	Arabisopsis thalia
35	46	59.0	1081	21	AAW50490	Arabisopsis thalia
36	46	59.0	1197	21	AAW50489	Arabisopsis thalia
37	44	56.4	343	16	AAW82455	Hepatitis GB virus
38	44	56.4	343	21	AAW09478	Hepatitis GB virus
39	44	56.4	2972	21	AAW09265	Hepatitis GB virus
40	44	56.4	3163	16	AAW94347	Hepatitis GB virus
41	43	55.1	16	18	AAW32456	Mycobacterium tube
42	43	55.1	16	19	AAW81694	M. tuberculosis im
43	43	55.1	16	20	AAV39124	M. tuberculosis an
44	43	55.1	454	21	AAV74537	Neisseria gonorrhoe
45	43	55.1	454	21	AAV74538	Neisseria meningit

## ALIGNMENTS

RESULT 1  
AAV03709  
ID AAV03709 standard; Protein; 16 AA.  
XX AC AAV03709;

XX 07-JUN-1999 (first entry)

DE M. tuberculosis LHP polypeptide antigenic fragment.

XX ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
KW immune response.

XX Mycobacterium tuberculosis.

XX WO9904005-A1.

XX 28-JAN-1999.

XX 16-JUL-1998; 98WO-IB01091.

XX 16-JUL-1997; 97US-0052631.

XX (INSP ) INST PASTEUR.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX WPI; 1999-132249/11.

XX New nucleic acid containing regulator and LHP gene of Mycobacterium  
PT tuberculosis - useful in vaccines, for diagnosis, and for expression  
PT of heterologous proteins

XX PS Claim 21; Page 65; 88pp; English.

XX CC The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX SQ Sequence 16 AA;

Query Match 100.0%; Score 78; DB 20; Length 16;

Best Local Similarity 100.0%; Pred. No. 4.4e-07;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16

DB 1 gdlktqdqvstags 16

RESULT 2

AAAY03706

ID AAY03706 standard; Protein; 49 AA.

XX AC AAY03706;

XX DT 07-JUN-1999 (first entry)

XX DE M. tuberculosis LHP polypeptide antigenic fragment.

XX KW ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;

XX KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;

XX KW immune response.

XX OS Mycobacterium tuberculosis.

XX PN WO9904005-A1.

XX PD 28-JAN-1999.

XX PF 16-JUL-1998; 98WO-IB01091.

XX PR 16-JUL-1997; 97US-0052631.

XX PA (INSP ) INST PASTEUR.

XX PA (STAT-) STATENS SERUM INST.

XX PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX WPI; 1999-132249/11.

XX New nucleic acid containing regulator and LHP gene of Mycobacterium

PT tuberculosis - useful in vaccines, for diagnosis, and for expression

PT of heterologous proteins

XX PS Claim 21; Page 64; 88pp; English.

XX CC The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX SQ Sequence 49 AA;

Query Match 100.0%; Score 78; DB 20; Length 49;

Best Local Similarity 100.0%; Pred. No. 1.5e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16

DB 23 gdlktqdqvstags 38

RESULT 3

AAW32454

ID AAW32454 standard; Protein; 80 AA.

XX AC AAW32454;

XX DT 09-JAN-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen Tb38-1F3.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

XX KW skin testing; M.tuberculosis.

XX OS Mycobacterium tuberculosis.

XX PN WO9709428-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US14674.

XX PR 12-JUL-1996; 96US-0680574.

XX PR 01-SEP-1995; 95US-0523436.

XX PR 22-SEP-1995; 95US-0533634.

XX PR 22-MAR-1996; 96US-0620874.

XX PR 05-JUN-1996; 96US-0659683.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

XX PI Twardzik DR, Vedvick TH;

XX DR WPI; 1997-192903/17.

XX DR N-PSDB; AAT91526.

XX PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are

PT useful in vaccines for prevention or treatment of tuberculosis, also

PT for diagnosis

XX PS Example 3; Page 149; 168pp; English.

XX CC A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or

CC its variant differing only in conservative substitutions and/or

CC modifications). The present sequence represents a M.tuberculosis

CC antigen, Tb38-1F3. The immunogenic protein, and fusion proteins

CC containing one or more of the proteins or one of the proteins plus

CC ESAT-6, are useful in vaccines, preferably when formulated with a

CC non-specific adjuvant, to induce an immune response against  
 XX M.tuberculosis (for treatment or prevention).  
 XX  
 SQ Sequence 80 AA;

Query Match 100.0%; Score 78; DB 18; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GDLKTOIDQVESTAGS 16  
 |||||  
 Db 3 gdlktqldqvstags 18

RESULT 4  
 AAW32386  
 ID AAW32386 standard; Protein; 80 AA.

XX AC AAW32386;

XX DT 13-JAN-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen Tb38-Ir3.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

XX KW skin testing; M.tuberculosis.

XX OS Mycobacterium tuberculosis.

XX PN W09709429-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US14675.

XX PR 12-JUL-1996; 96US-0680573.

XX PR 01-SEP-1995; 95US-0523435.

XX PR 22-SEP-1995; 95US-0532136.

XX PR 23-MAR-1996; 96US-0620280.

XX PR 05-JUN-1996; 96US-0658800.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

XX PI Twardzik DR, Vedvick TH;

XX PT WPI; 1997-192904/17.

XX DR N-PSDB; AAT91460.

XX PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 - useful for diagnosis of M. tuberculosis infection

XX PS Example 3; Page 163; 190pp; English.

XX CC A new immunogenic polypeptide has been developed comprising an  
 immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 its variant differing only in conservative substitutions and/or  
 modifications). The present sequence represents a M.tuberculosis  
 antigen, Tb38-Ir3. The immunogenic polypeptide can be used to diagnose  
 M.tuberculosis infection by forming complexes with specific  
 antibodies in the sample. Fragments of DNA encoding the immunogenic  
 polypeptide can be used as diagnostic primers or probes and agents  
 that bind to the antigen, especially monoclonal antibodies or  
 equivalent polyclonal antibodies, are also used for diagnosis.

XX SQ Sequence 80 AA;

Query Match 100.0%; Score 78; DB 18; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GDLKTOIDQVESTAGS 16  
 |||||  
 Db 3 gdlktqldqvstags 18

RESULT 5

AAW81707  
 ID AAW81707 standard; Protein; 80 AA.

XX AC AAW81707;

XX DT 27-JAN-1999 (first entry)

XX DE M. tuberculosis immunogenic polypeptide Tb38-Ir3.

XX KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.

XX OS Mycobacterium tuberculosis.

XX PN W09816646-A2.

XX PD 23-APR-1998.

XX PF 07-OCT-1997; 97WO-US18293.

XX PR 13-MAR-1997; 97US-0818112.

XX PR 11-OCT-1996; 96US-0730510.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX DR WPI; 1998-261042/23.

XX DR N-PSDB; AAV64509.

XX PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 to develop products for the detection of M. tuberculosis infection  
 and for diagnosis, treatment and prevention of tuberculosis

XX PS Example 3B; Page 139-140; 230pp; English.

XX CC This sequence represents an immunogenic portion of a soluble  
 Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 for inducing protective immunity against tuberculosis (TB). This  
 sequence can be formulated into vaccines and/or pharmaceutical  
 compositions for immunising against M. tuberculosis infection or may  
 be used for the diagnosis of tuberculosis.

XX SQ Sequence 80 AA;

Query Match 100.0%; Score 78; DB 19; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GDLKTOIDQVESTAGS 16  
 |||||  
 Db 3 gdlktqldqvstags 18

RESULT 6

AAW64340  
 ID AAW64340 standard; Protein; 80 AA.

XX AC AAW64340;

XX DT 09-NOV-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen Tb38-Ir3.

XX KW Tuberculosis; infection; diagnosis; antigen; Tb38-Ir3.

```

XX OS Mycobacterium tuberculosis strain H37Rv.
XX PN WO9816645-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18214.
XX PR 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1998-251292/22.
XX DR N-PSDB; AAV44400.
XX PR New isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT to develop products for the detection of M. tuberculosis infection
PT and diagnosis of tuberculosis
XX PS Example 3; Page 146; 250pp; English.
XX CC This polypeptide comprises Mycobacterium tuberculosis antigen
CC Tb38-1F3. A DNA molecule (see AAV44400) coding for Tb38-1F3 was
CC isolated from a M. tuberculosis strain H37Rv genomic library. The
CC invention relates to compositions and methods for diagnosing
CC tuberculosis. It provides polypeptides (see AAV44291-W64379)
CC comprising an antigenic portion of a soluble M. tuberculosis
CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
CC as well as DNA sequences encoding such polypeptides, recombinant
CC expression vectors and transformed or transfected host cells. Also
CC claimed are methods and diagnostic kits for detecting M.
CC tuberculosis infection in a patient using these polypeptides,
CC antibodies or oligonucleotide probes and primers, for the diagnosis
CC of tuberculosis.
XX SQ Sequence 80 AA;

Query Match 100.0%; Score 78; DB 19; Length 80;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16
Db 3 gdlktqidqvstags 18

RESULT 7
AAV39137
ID AAV39137 standard; Protein; 80 AA.
XX AC AAV39137;
XX DT 05-NOV-1999 (first entry)
XX DE M. tuberculosis antigen Tb38-1F3 amino acid sequence.
XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX OS Mycobacterium tuberculosis.
XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US03268.

```

```

XX 05-MAY-1998; 98US-0072967.
XX 18-FEB-1998; 98US-0025197.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1999-527409/44.
XX DR N-PSDB; AAZ19310.
XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions
XX PS Example 3; Page 134-135; 299pp; English.
XX CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAZ39083 to
CC AAZ39225 are used in the exemplification of the present invention.
XX SQ Sequence 80 AA;

Query Match 100.0%; Score 78; DB 20; Length 80;
Best Local Similarity 100.0%; Pred. No. 2.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16
Db 3 gdlktqidqvstags 18

RESULT 8
AAV38994
ID AAV38994 standard; Protein; 80 AA.
XX AC AAV38994;
XX DT 05-NOV-1999 (first entry)
XX DE M. tuberculosis recombinant antigen protein Tb38-1F3.
XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX OS Mycobacterium tuberculosis.
XX PN WO9942118-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US03265.
XX PR 05-MAY-1998; 98US-0072596.
XX PR 18-FEB-1998; 98US-0024753.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1999-527416/44.
XX DR N-PSDB; AAZ19098.
XX PF

```



PT New polypeptide comprising antigenic portions of M. tuberculosis  
XX  
PS Example 3; Page 180; 323pp; English.  
XX

CC This invention describes novel recombinant antigens and their encoding  
CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
CC polypeptides are useful for detecting M. tuberculosis infection in a  
CC biological sample by detecting antibodies which bind with the  
CC polypeptides, and are useful as vaccines for immunizing against  
CC M. tuberculosis infection. The new detection methods are needed as  
CC current vaccination strategies do not provide 100% immunity.  
XX  
SQ Sequence 80 AA;

Query Match 100.0%; Score 78; DB 20; Length 80;  
Best Local Similarity 100.0%; Pred. No. 2.5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GDLKTQIDQVESTAGS 16  
Db 3 gdlktqidqvstags 18  
|||||

RESULT 9  
AAW32444  
ID AAW32444 standard; Protein; 95 AA.  
XX  
AC AAW32444;  
XX  
DT 09-JAN-1998 (first entry)  
XX  
DE Mycobacterium tuberculosis antigen Tb38-1.  
XX  
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
XX skin testing; M.tuberculosis.  
OS Mycobacterium tuberculosis.  
XX  
PN WO9709428-A2.  
XX  
PD 13-MAR-1997.  
XX  
PF 30-AUG-1996; 96WO-US14674.  
XX  
PR 12-JUL-1996; 96US-0680574.  
XX  
PR 01-SEP-1995; 95US-0523436.  
XX  
PR 22-SEP-1995; 95US-0533634.  
XX  
PR 22-MAR-1996; 96US-0620874.  
XX  
PR 05-JUN-1996; 96US-0659683.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
XX Twardzik DR, Vedvick TH;  
XX  
XX WPI; 1997-192903/17.  
XX N-PSDB; AAT91509.  
XX

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
XX useful in vaccines for prevention or treatment of tuberculosis, also  
XX for diagnosis  
XX  
PS Example 3; Page 124; 168pp; English.  
XX  
CC A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb38-1. The immunogenic protein, and fusion proteins  
CC containing one or more of the proteins or one of the proteins plus  
CC ESAT-6, are useful in vaccines, preferably when formulated with a  
CC non-specific adjuvant, to induce an immune response against

Query Match 100.0%; Score 78; DB 20; Length 80;  
Best Local Similarity 100.0%; Pred. No. 2.5e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GDLKTQIDQVESTAGS 16  
Db 3 gdlktqidqvstags 18  
|||||

CC M.tuberculosis (for treatment or prevention).  
XX  
SQ Sequence 95 AA;

Query Match 100.0%; Score 78; DB 18; Length 95;  
Best Local Similarity 100.0%; Pred. No. 3e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GDLKTQIDQVESTAGS 16  
Db 18 gdlktqidqvstags 33  
|||||

RESULT 10  
AAW32376  
ID AAW32376 standard; Protein; 95 AA.  
XX  
AC AAW32376;  
XX  
DT 13-JAN-1998 (first entry)  
XX  
DE Mycobacterium tuberculosis antigen Tb38-1.  
XX  
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
XX skin testing; M.tuberculosis.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO9709429-A2.  
XX  
PD 13-MAR-1997.  
XX  
PF 30-AUG-1996; 96WO-US14675.  
XX  
PR 12-JUL-1996; 96US-0680573.  
XX  
PR 01-SEP-1995; 95US-0523435.  
XX  
PR 22-SEP-1995; 95US-0532136.  
XX  
PR 22-MAR-1996; 96US-0620280.  
XX  
PR 05-JUN-1996; 96US-0658800.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
XX Twardzik DR, Vedvick TH;  
XX  
XX WPI; 1997-192904/17.  
XX N-PSDB; AAT91445.  
XX  
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
XX - useful for diagnosis of M. tuberculosis infection  
XX  
PS Example 3; Page 136; 190pp; English.  
XX  
CC A new immunogenic polypeptide has been developed comprising an  
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose  
CC M.tuberculosis infection by forming complexes with specific  
CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
CC polypeptide can be used as diagnostic primers or probes and agents  
CC that bind to the antigen, especially monoclonal antibodies or  
CC equivalent polyclonal antibodies, are also used for diagnosis.  
XX  
SQ Sequence 95 AA;

Query Match 100.0%; Score 78; DB 18; Length 95;  
Best Local Similarity 100.0%; Pred. No. 3e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GDLKTQIDQVESTAGS 16

Db 18 gdlktqidqvestags 33  
 |||

## RESULT 11

AAW81747  
 ID AAW81747 standard; Protein; 95 AA.

AC AAW81747;

DT 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide Tb38-1.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

OS WO9816646-A2.

PN 23-APR-1998.

PF 07-OCT-1997; 97WO-US18293.

PR 13-MAR-1997; 97US-0818112.

PR 11-OCT-1996; 96US-0730510.

PA (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Read SG, Skeiky YAW, Twardzik DR, Vedvick TS;

DR WPI; 1998-261042/23.

DR N-PSDB; AAV64491.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis

PS Example 3b; Page 117; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This sequence  
 CC can be formulated into vaccines and/or pharmaceutical compositions for  
 CC immunising against M. tuberculosis infection or may be used for the  
 CC diagnosis of tuberculosis.

XX SQ Sequence 95 AA;

## Query Match

Best Local Similarity 100.0%; Score 78; DB 19; Length 95;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVESTAGS 16

Db 18 gdlktqidqvestags 33

## RESULT 12

AAW64321  
 ID AAW64321 standard; Peptide; 95 AA.

AC AAW64321;

DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb38-1 peptide.

XX Tuberculosis; infection; diagnosis; antigen; Tb38-1.

XX

OS Mycobacterium tuberculosis strain H37Rv.

PN WO9816645-A2.

PD 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

PR 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Read SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

DR N-PSDB; AAV44384.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and diagnosis of tuberculosis

PS Example 3; Page 123; 250pp; English.

XX This is an antigenic portion of Mycobacterium tuberculosis antigen  
 CC Tb38-1. A DNA sequence (see AAV44384) coding for antigen Tb38-1 was  
 CC isolated from a M. tuberculosis strain H37Rv expression library  
 CC using sera from patients having pulmonary or pleural tuberculosis.  
 CC The invention relates to compositions and methods for diagnosing  
 CC tuberculosis. It provides polypeptides (see AAW64291-W64379)  
 CC comprising an antigenic portion of a soluble M. tuberculosis  
 CC antigen, or an immunogenic portion of an M. tuberculosis antigen,  
 CC as well as DNA sequences encoding such polypeptides, recombinant  
 CC expression vectors and transformed or transfected host cells. Also  
 CC claimed are methods and diagnostic kits for detecting M.  
 CC tuberculosis infection in a patient using these polypeptides,  
 CC antibodies or oligonucleotide probes and primers, for the diagnosis  
 CC of tuberculosis.

XX SQ Sequence 95 AA;

## Query Match

Best Local Similarity 100.0%; Score 78; DB 19; Length 95;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVESTAGS 16

Db 18 gdlktqidqvestags 33

## RESULT 13

AAW32097

ID AAW32097 standard; Protein; 95 AA.

XX AAW32097;

DT 17-JAN-2000 (first entry)

XX Mycobacterium tuberculosis antigen Tb38-1.

XX Tuberculosis; antigen; fusion protein; Tb38-1; diagnosis; therapy;  
 KW vaccine; immunogen.

XX Mycobacterium tuberculosis.

XX WO9951748-A2.

XX 14-OCT-1999.

XX 07-APR-1999; 99WO-US07717.

XX

PR 07-APR-1998; 98US-0056556.  
 PR 30-DEC-1998; 98US-0223040.  
 PA (CORI-) CORIXA CORP.  
 PI Skelky YAW, Alderson M, Campos-Neto A;  
 DR WPI; 1999-601610/51.  
 XX New fusion proteins useful for diagnosis, prevention and treatment of  
 PT tuberculosis -  
 XX Claim 1; Fig 4D; 83pp; English.  
 CC This sequence represents the Mycobacterium tuberculosis antigen  
 CC Tb38-1. The invention provides fusion proteins (see AAY32059-71)  
 CC containing at least 2 M. tuberculosis antigens such as Tb38-1, e.g.  
 CC TbF-2 (see AAY32063) and a TbH9-Tb38-1 fusion. The new fusion  
 CC proteins are useful as vaccines for preventing tuberculosis  
 CC (claimed), for diagnosis (via in vitro assays or intradermal skin  
 CC tests for detection of anti-M. tuberculosis antibodies), monitoring  
 CC of disease progression, and treatment of tuberculosis. They are  
 CC more effective immunogens than mixtures of the individual protein  
 CC components.  
 XX Sequence 95 AA;  
 SQ

Query Match 100.0%; Score 78; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDLKTDIDQVESTAGS 16  
 |||||  
 Db 18 gdlktqldqvstags 33

RESULT 14  
 AAY39118  
 ID AAY39118 standard; Protein; 95 AA.  
 XX  
 AC AAY39118;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis antigen Tb38-1 amino acid sequence.  
 XX  
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9942076-A2.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99WO-US03268.  
 XX  
 PR 05-MAY-1998; 98US-0072967.  
 PR 18-FEB-1998; 98US-0025197.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1999-527409/44.  
 XX  
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX  
 PS Example 3; Page 113; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AAY19249 to AAY19460 and AAY39083 to  
 CC AAY39225 are used in the exemplification of the present invention.  
 XX  
 SQ Sequence 95 AA;  
 Query Match 100.0%; Score 78; DB 20; Length 95;  
 Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GDLKTDIDQVESTAGS 16  
 |||||  
 Db 18 gdlktqldqvstags 33

RESULT 15  
 AAY38981  
 ID AAY38981 standard; Protein; 95 AA.  
 XX  
 AC AAY38981;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis recombinant antigen protein Tb38-1.  
 XX  
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO9942118-A2.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99WO-US03265.  
 XX  
 PR 05-MAY-1998; 98US-0072596.  
 PR 18-FEB-1998; 98US-0024753.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1999-527416/44.  
 DR N-PSDB; AAY19082.  
 XX  
 PT New polypeptide comprising antigenic portions of M. tuberculosis  
 XX  
 PS Example 3; Page 159; 323pp; English.  
 CC This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.  
 XX  
 SQ Sequence 95 AA;  
 Query Match 100.0%; Score 78; DB 20; Length 95;

Best Local Similarity 100.0%; Pred. No. 3e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GDLKTCIDQVESTAGS 16  
Db 18 gdlktqidqvestags 33

Search completed: July 5, 2001, 11:45:43  
Job time: 272 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:46:59 ; Search time 61.79 Seconds  
(without alignments)  
5.216 Million cell updates/sec

Title: US-09-462-480-9

Perfect score: 78

Sequence: 1 GDLKTQIDQVESTAGS 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	53.8	2756	1	US-08-375-709-11
2	42	53.8	2756	1	US-08-752-929-11
3	42	53.8	2756	4	US-09-090-793-7
4	41	52.6	247	3	US-08-969-644-18
5	41	52.6	247	4	US-08-444-189-18
6	39	50.0	396	1	US-08-430-024-2
7	39	50.0	396	1	US-08-782-009-2
8	39	50.0	396	3	US-09-017-302-2
9	39	50.0	577	2	US-08-756-317-13
10	39	50.0	600	6	5240706-1
11	39	50.0	984	1	US-08-242-932-2
12	39	50.0	984	1	US-08-714-481-2
13	39	50.0	984	5	PCT-US95-06111-2
14	38	48.7	278	1	US-08-188-582-9
15	38	48.7	278	1	US-08-646-715-9
16	38	48.7	1832	4	US-09-335-409-4
17	36	46.2	176	1	US-08-726-525-2
18	36	46.2	176	2	US-08-487-942-2
19	36	46.2	176	2	US-08-726-036A-2
20	36	46.2	437	2	US-08-781-560-1
21	36	46.2	437	2	US-08-781-560-3
22	36	46.2	437	3	US-08-792-014-3
23	36	46.2	437	4	US-09-443-948-3
24	36	46.2	454	3	US-08-446-100-26
25	36	46.2	454	3	US-08-446-100-27
26	36	46.2	454	3	US-08-446-100-28
27	36	46.2	454	3	US-08-446-100-29

Sequence 30, Appl  
Sequence 31, Appl  
Sequence 5, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 10, Appl  
Sequence 29, Appl  
Sequence 10, Appl  
Sequence 6, Appl  
Sequence 8, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 6, Appl  
Sequence 50, Appl  
Sequence 23, Appl

28 36 46.2 454 3 US-08-446-100-30  
29 36 46.2 454 3 US-08-446-100-31  
30 36 46.2 737 1 US-08-119-361-5  
31 36 46.2 737 3 US-08-336-308A-4  
32 36 46.2 737 4 US-08-822-324-4  
33 36 46.2 893 1 US-07-977-434-4  
34 36 46.2 893 1 US-08-458-819-4  
35 36 46.2 893 4 US-09-105-697-10  
36 36 46.2 893 5 PCT-US91-07035-4  
37 36 46.2 1687 2 US-08-570-311-29  
38 36 46.2 1704 3 US-08-336-308A-10  
39 36 46.2 1704 4 US-08-822-324-6  
40 36 46.2 4550 2 US-08-804-227C-8  
41 36 46.2 4550 2 US-08-804-198-2  
42 36 46.2 4551 4 US-09-320-878-1  
43 35.5 45.5 578 4 US-09-066-046-6  
44 35.5 45.5 578 4 US-08-975-762-50  
45 35 44.9 281 1 US-07-941-523-23

#### ALIGNMENTS

RESULT 1  
US-08-375-709-11  
; Sequence 11, Application US/08375709  
; Patent No. 5683898  
; GENERAL INFORMATION:  
; APPLICANT: YAZAWA, Kazunaga  
; APPLICANT: YAMADA, Akiko  
; APPLICANT: KATO, Seishi  
; APPLICANT: KONDO, Kiyosi  
; TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid  
; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of  
; TITLE OF INVENTION: Eicosapentaenoic Acid  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/375,709  
; FILING DATE: 20-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/178,251  
; FILING DATE: 14-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-147945  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WEGNER, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2756 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-375-709-11

Query Match 53.8%; Score 42; DB 1; Length 2756;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16

Db 397 GSVKSQIGHTKTSTAGT 412

RESULT 2

US-08-752-929-11  
; Sequence 11, Application US/08752929  
; Patent No. 5798259  
; GENERAL INFORMATION:  
; APPLICANT: YAZAWA, Kazunaga  
; APPLICANT: YAMADA, Akiko  
; APPLICANT: KATO, Seishi  
; APPLICANT: KONDO, Kiyosi  
; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing  
; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic  
; TITLE OF INVENTION: Acid  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/752.929  
; FILING DATE: 20-NOV-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/375,709  
; FILING DATE: 20-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/178,251  
; FILING DATE: 14-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 4-147945  
; FILING DATE: 15-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WEGNER, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2756 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-752-929-11

Query Match 53.8%; Score 42; DB 1; Length 2756;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16

Db 397 GSVKSQIGHTKTSTAGT 412

RESULT 3

US-09-090-793-7  
; Sequence 7, Application US/09090793  
; Patent No. 6140486  
; GENERAL INFORMATION:  
; APPLICANT: Calgene, LLC  
; TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
; TITLE OF INVENTION: of polyketide-like synthesis genes in plants  
; FILE REFERENCE: CGNE.131.01US  
; CURRENT APPLICATION NUMBER: US/09/090.793  
; CURRENT FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048.650  
; EARLIER FILING DATE: 1997-06-04  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 2756  
; TYPE: PRT  
; ORGANISM: Shewanella putrefaciens  
US-09-090-793-7

Query Match 53.8%; Score 42; DB 4; Length 2756;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16

Db 397 GSVKSQIGHTKTSTAGT 412

RESULT 4

US-08-969-644-18  
; Sequence 18, Application US/08969644  
; Patent No. 6096519  
; GENERAL INFORMATION:  
; APPLICANT: Ratti, Giulio  
; APPLICANT: Comanducci, Maurizio  
; APPLICANT: Tecce, Mario F.  
; APPLICANT: Giuliani, Marzia M.  
; TITLE OF INVENTION: PCTD PLASMID ISOLATED FROM CHLAMYDIA  
; TITLE OF INVENTION: TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED BY  
; TITLE OF INVENTION: THEM; RECOMBINANT PLASMIDS FOR THE EXPRESSION OF SAID  
; TITLE OF INVENTION: GENES IN HETEROLOGOUS SYSTEMS, PREPARATION OF SAID  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 301 N. Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22046-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/969.644  
; FILING DATE: 13-NOV-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/467,152  
; FILING DATE:  
; APPLICATION NUMBER: US/07/661,820  
; FILING DATE:  
; APPLICATION NUMBER: IT MI 91A000314  
; FILING DATE: 07-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svensson, Leonard R.  
; REGISTRATION NUMBER: 30,330

; REFERENCE/DOCKET NUMBER: 1267-202P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 247 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-969-644-18

Query Match 52.6%; Score 41; DB 3; Length 247;  
Best Local Similarity 46.7%; Pred. No. 6.1;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAG 15  
|||||:|:|  
Db 160 GDLTKVDVIGKVC 174

RESULT 5  
US-08-444-189-18  
; Sequence 18, Application US/08444189  
; Patent No. 6110705  
; GENERAL INFORMATION:  
; APPLICANT: Ratti, Giulio  
; APPLICANT: Comanducci, Maurizio  
; APPLICANT: Tecce, Mario F.  
; APPLICANT: Giuliani, Marzia M.  
; TITLE OF INVENTION: PCTD PLASMTID ISOLATED FROM CHLAMYDIA  
; TITLE OF INVENTION: TRACHOMATIS SEROTYPE D, ITS GENES AND PROTEINS ENCODED BY  
; TITLE OF INVENTION: THEM; RECOMBINANT PLASMTIDS FOR THE EXPRESSION OF SAID  
; TITLE OF INVENTION: GENES IN HETEROLOGOUS SYSTEMS, PREPARATION OF SAID  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
; STREET: 301 N. Washington Street  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22046-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/444,189  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/180,528  
; FILING DATE:  
; APPLICATION NUMBER: US/07/991,712  
; FILING DATE:  
; APPLICATION NUMBER: US/07/661,820  
; FILING DATE:  
; APPLICATION NUMBER: IT MI 91A000314  
; FILING DATE: 07-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Svensson, Leonard R.  
; REGISTRATION NUMBER: 30,330  
; REFERENCE/DOCKET NUMBER: 1267-202P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 247 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-444-189-18

Query Match 52.6%; Score 41; DB 4; Length 247;  
Best Local Similarity 46.7%; Pred. No. 6.1;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAG 15  
|||||:|:|  
Db 160 GDLTKVDVIGKVC 174

RESULT 6  
US-08-430-024-2  
; Sequence 2, Application US/08430024  
; Patent No. 5644030  
; GENERAL INFORMATION:  
; APPLICANT: Faulmann, Ervin L.  
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF  
; TITLE OF INVENTION: AN Iga BINDING PROTEIN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/430,024  
; FILING DATE: 27-APR-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/677,209  
; FILING DATE: 29-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: BL-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 396 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-430-024-2

Query Match 50.0%; Score 39; DB 1; Length 396;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 DLKTQIDQVESTAG 15  
|:|:|:|:|  
Db 84 DIRKAVEPVEKTAG 97

RESULT 7  
US-08-782-009-2  
; Sequence 2, Application US/08782009  
; Patent No. 5714334  
; GENERAL INFORMATION:

```

RESULT      8
US-09-017-302-2
; Sequence 2, Application US/09017302
; Patent No. 6075128
; GENERAL INFORMATION:
; APPLICANT: Faulmann, Ervin L
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF
; SEQUENCE 2, APPLICATION: AN IGA BINDING PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/017.302

```

APPLICANT: Kishore, Ganesh M.  
APPLICANT: Mitsky, Timothy A.  
APPLICANT: Stark, David M.  
TITLE OF INVENTION: Improved Rhodospirillum Rubrum  
TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/756,317  
FILING DATE: 25-NOV-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,693  
FILING DATE: 29-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Patterson, Melinda L.  
REGISTRATION NUMBER: 33,062  
REFERENCE/DOCKET NUMBER: MOBT:008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 787-1400  
TELEFAX: (713) 787-1440  
INFORMATION FOR SEQ ID NO: 13:



; SEQUENCE CHARACTERISTICS:  
; LENGTH: 577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-756-317-13

Query Match 50.0%; Score 39; DB 2; Length 577;  
Best Local Similarity 43.8%; Pred. No. 37;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 GDLKTDQIDQVESTAGS 16  
||:| : : : ||||  
Db 521 GDIETWGRKAKETAGS 536

RESULT 10  
5240706-1  
; Patent No. 5240706  
; APPLICANT: FAULDS, DARYL  
; TITLE OF INVENTION: INTRANASSAL ADMINISTRATION OF MYCOPLASMA  
; HYOPNEUMONIAE ANTIGEN  
; NUMBER OF SEQUENCES: 22  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/334,586  
; FILING DATE: 07-APR-1989  
; SEQ ID NO:1;  
; LENGTH: 600  
5240706-1

Query Match 50.0%; Score 39; DB 6; Length 600;  
Best Local Similarity 46.7%; Pred. No. 38;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 DLKTDQIDQVESTAGS 16  
|| :||:| : |  
Db 555 ELKLDQIEAAQS 569

RESULT 11  
US-08-242-932-2  
; Sequence 2, Application US/08242932  
; Patent No. 5595740  
; GENERAL INFORMATION:  
; APPLICANT: Brady, L. Jeannine  
; TITLE OF INVENTION: Cloning of No. 5595740-IgA Fc Binding Forms of  
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/242,932  
; FILING DATE: 16-MAY-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: UF142  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-372-5800

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 984 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-242-932-2

Query Match 50.0%; Score 39; DB 1; Length 984;  
Best Local Similarity 50.0%; Pred. No. 68;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 DLKTDQIDQVESTAG 15  
|:| :| :|||  
Db 47 DIKKAPEPVEKTAG 60

RESULT 12  
US-08-714-481-2  
; Sequence 2, Application US/08714481  
; Patent No. 5766606  
; GENERAL INFORMATION:  
; APPLICANT: Brady, L. Jeannine  
; TITLE OF INVENTION: Cloning of No. 5766606-IgA Fc Binding Forms of  
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/714,481  
; FILING DATE: 16-SEP-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/242,932  
; FILING DATE: 16-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: UF142  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 984 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-714-481-2

Query Match 50.0%; Score 39; DB 1; Length 984;  
Best Local Similarity 50.0%; Pred. No. 68;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 DLKTDQIDQVESTAG 15  
|:| :| :|||  
Db 47 DIKKAPEPVEKTAG 60

RESULT 13  
PCT-US95-061111-2



```

; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,882
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-646-715-9

```

```

Query Match      48.7%; Score 38; DB 1; Length 278;
Best Local Similarity 43.8%; Pred. No. 24;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 GDLKTOIDQVESTAGS 16
Db 221 GDLKMEYDSDAAVGS 236

```

Search completed: July 5, 2001, 11:47:00  
Job time: 314 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:48:34 ; Search time 79.63 Seconds  
(without alignments)  
15.306 Million cell updates/sec

Title: US-09-462-480-9  
Perfect score: 78  
Sequence: 1 GDLKTQIDQVESTAGS 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	100	2 H70802	hypothetical prote
2	46	59.0	1163	2 F84669	probable chromosom
3	43.5	55.8	630	2 T38023	probable transcrip
4	43	55.1	1068	2 E81965	probable outer mem
5	43	55.1	1082	2 H81020	serotype-1-specifi
6	42	53.8	822	2 JC4076	dextranase (EC 3.2
7	42	53.8	2756	2 T30183	hypothetical prote
8	41	52.6	247	2 S01924	hypothetical prote
9	41	52.6	247	2 F37386	hypothetical prote
10	41	52.6	497	2 B71677	histidine kinase s
11	41	52.6	533	2 G72593	hypothetical prote
12	40	51.3	388	2 H86843	chorismate synthas
13	40	51.3	442	2 T18917	hypothetical prote
14	40	51.3	447	2 S44809	F44B9.8 protein -
15	40	51.3	1019	2 T11560	pol polyprotein -
16	40	51.3	1096	2 T16875	hypothetical prote
17	39	50.0	523	2 S53945	hypothetical prote
18	39	50.0	774	2 T34496	hypothetical prote
19	39	50.0	1134	2 A60234	IgA Fc receptor pr
20	39	50.0	1164	1 FCSOAG	IgA Fc receptor pr
21	39	50.0	1390	2 T31353	polyprotein - Arab
22	38.5	49.4	129	2 S03478	T-cell receptor al
23	38	48.7	128	2 C69295	transcription repr
24	38	48.7	278	2 A49067	transcription init
25	38	48.7	334	2 T03556	probable D-ribose-
26	38	48.7	414	2 T15947	hypothetical prote
27	38	48.7	550	2 T34771	signal recognition
28	38	48.7	585	2 T11997	acetohydroxyacid s
29	38	48.7	767	2 D70131	conserved hypothet

30 38 48.7 816 2 S64439 hypothetical prote  
31 38 48.7 1733 2 D70887 probable polyketid  
32 38 48.7 2077 1 WZBE24 240K tegument prot  
33 38 48.7 2077 2 T43991 large tegument pro  
34 38 48.7 2077 2 T44178 probable aspartate  
35 37.5 48.1 312 2 C72547 probable aspartate  
36 37.5 48.1 365 2 T25296 hypothetical prote  
37 37 47.4 102 2 B31512 parathymosin - rat  
38 37 47.4 106 2 S20422 zinc-binding prote  
39 37 47.4 132 1 JVBPF1 DNA-packaging prot  
40 37 47.4 172 2 E81897 probable host-nucl  
41 37 47.4 298 2 T03990 probable transcrip  
42 37 47.4 328 2 E84243 transcription init  
43 37 47.4 332 2 T42370 probable glucan 1.  
44 37 47.4 343 2 T32232 hypothetical prote  
45 37 47.4 352 1 A34441 farnesyl-pyrophosp

## ALIGNMENTS

RESULT 1  
H70802  
hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70802  
R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: H70802  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-100 <COL>  
A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL17966.1; PID:g296  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3874

Query Match 100.0%; Score 78; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 2.5e+06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 GDLKTQIDQVESTAGS 16  
|||||  
DB 23 GDLKTQIDQVESTAGS 38

RESULT 2  
F84669  
probable chromosome associated protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: F84669  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: F84669  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1163 <STO>  
A:Cross-references: GB:AE002093; NID:g6598752; PIDN:AAD26882.2; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g27170  
A:Map position: 2

C:Superfamily: hypothetical protein YJL074c

Query Match 59.0%; Score 46; DB 2; Length 1163;  
Best Local Similarity 61.5%; Pred. No. 13;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVEST 13  
|||||  
DB 710 GDITRIDQVRSS 722

RESULT 3

T38023  
probable transcription regulator - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 07-Dec-1999  
C:Accession: T38023  
R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1997

A:Reference number: 221763  
A:Accession: T38023  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-630 <MUR>  
A:Cross-references: EMBL:Z98598; PIDN:CAB11234.1; GSPDB:GNO0066; SPDB:SPAC1B3.05  
A:Experimental source: strain 972h-; cosmid c1B3  
C:Genetics:  
A:Gene: SPDB:SPAC1B3.05  
A:Map position: 1  
A:Introns: 75/3

Query Match 55.8%; Score 43.5; DB 2; Length 630;  
Best Local Similarity 47.8%; Pred. No. 17;  
Matches 11; Conservative 3; Mismatches 2; Indels 7; Gaps 1;

QY 1 GDLKTQI-----DOVESTAGS 16  
|||||  
DB 35 GDLKTQIKKQLRLRDIKTWASS 57

RESULT 4

E81965  
probable outer membrane peptidase (EC 3.4.21.-) NMA0478 [imported] - Neisseria meningitidis

C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: E81965  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556

A:Accession: E81965  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1068 <PAR>  
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83773.1; PID:g737922  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0478  
C:Keywords: hydrolase; serine proteinase

Query Match 55.1%; Score 43; DB 2; Length 1068;  
Best Local Similarity 56.2%; Pred. No. 38;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16  
|||||  
DB 654 GGLLASLDSVETKTAGS 669

RESULT 5

H81020  
serotype-1-specific antigen, probable NMB1969 [imported] - Neisseria meningitidis (st

C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: H81020  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755

A:Accession: H81020  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1082 <ET>  
A:Cross-references: GB:AE002545; GB:AE002098; NID:g727229; PIDN:AAF42298.1; PID:g722  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1969

Query Match 55.1%; Score 43; DB 2; Length 1082;  
Best Local Similarity 56.2%; Pred. No. 38;  
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16  
|||||  
DB 668 GGLLASLDSVETKTAGS 683

RESULT 6

JC4076

dextranase (EC 3.2.1.11) precursor - Streptococcus salivarius  
N:Alternate names: 1,6-alpha-glucanhydrolase  
C:Species: Streptococcus salivarius  
C:Date: 21-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 15-Oct-1999  
C:Accession: JC4076  
R:Ohnishi, Y.; Kubo, S.; Ono, Y.; Nozaki, M.; Gonda, Y.; Okano, H.; Matsuya, T.; Mats  
Gene 156, 93-96, 1995  
A:Title: Cloning and sequencing of the gene coding for dextranase from Streptococcus  
A:Reference number: JC4076; MUID:95255675  
A:Accession: JC4076  
A:Molecule type: DNA  
A:Residues: 1-822 <OHN>

A:Cross-references: DDBJ:D29644; NID:g961504; PIDN:BA06127.1; PID:dl006681; PID:g961  
A:Experimental source: M-33  
A:Note: The authors' translation from residues 2303 to 2401 differs considerably from  
C:Comment: This enzyme hydrolyzes the 1,6-alpha-glucosidic linkages of glucan and re  
C:Genetics:  
A:Gene: dex

C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-38/Domain: signal sequence #status predicted <SIG>  
F:39-822/Product: dextranase #status predicted <MAT>

Query Match 53.8%; Score 42; DB 2; Length 822;  
Best Local Similarity 53.3%; Pred. No. 42;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAG 15  
|||||  
DB 334 GDLKARIDMVRNKTG 348

RESULT 7

T30183

hypothetical protein 5 - Shewanella sp.

C:Species: Shewanella sp.  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
C:Accession: T30183  
R:Takeyama, H.; Takeda, D.; Yazawa, K.; Yamada, A.; Matsunaga, T.

Microbiology 143, 2725-2731, 1997  
A:Title: Expression of the eicosapentaenoic acid synthesis gene cluster from *Shewanella*  
A:Reference number: Z20764; MUID:97419510  
A:Accession: T30183  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2756 <TAK>  
A:Cross-references: EMBL:U73935; NID:g2529413; PID:g2529418; PIDN:AAB81123.1  
A:Experimental source: strain SCRC-2738

Query Match 53.8%; Score 42; DB 2; Length 2756;  
Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAGS 16  
| : : : : :  
Db 397 GSVKSOIGTKSTAGT 412

RESULT 8

S01924 hypothetical protein 6 - Chlamydia trachomatis plasmids

N:Alternate names: Hypothetical protein P-10

C:Species: Chlamydia trachomatis

C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 08-Oct-1999

C:Accession: S01924; S00794

R:Comanducci, M.; Ricci, S.; Ratti, G.

Mol. Microbiol. 2, 531-538, 1988

A:Title: The structure of a plasmid of Chlamydia trachomatis believed to be required for

A:Reference number: S01920; MUID:89013895

A:Accession: S01924

A:Molecule type: DNA

A:Residues: 1-247 <COM>

A:Cross-references: EMBL:X07547; NID:g40730; PIDN:CAA30426.1; PID:g581014

A:Experimental source: plasmid pCHL1

A>Note: the authors translated the initiation codon GTG for residue 1 as Val

R:Hatt, C.; Ward, M.E.; Clarke, I.N.

Nucleic Acids Res. 16, 4053-4067, 1988

A:Title: Analysis of the entire nucleotide sequence of the cryptic plasmid of Chlamydia

A:Reference number: S00788; MUID:88233998

A:Accession: S00794

A:Molecule type: DNA

A:Residues: 175-247 <HAT>

A:Cross-references: EMBL:X06707

A:Experimental source: plasmid pLGV440

C:Genetics:

A:Genome: plasmid

A:Start codon: GTG

Query Match 52.6%; Score 41; DB 2; Length 247;  
Best Local Similarity 46.7%; Pred. No. 16;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAG 15  
| : : : : :  
Db 160 GDLKTKVDVICKVCG 174

RESULT 9

F37386

hypothetical protein 6 - Chlamydia trachomatis (serotype D) plasmid pCHL1

C:Species: Chlamydia trachomatis

C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 17-Feb-1994

C:Accession: F37386

R:Comanducci, M.; Ricci, S.; Cevenini, R.; Ratti, G.

Plasmid 23, 149-154, 1990

A:Title: Diversity of the Chlamydia trachomatis common plasmid in biovars with different

A:Reference number: A37386; MUID:90301796

A:Accession: F37386

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-247 <COM>  
A:Cross-references: GB:J03321  
A:Genetics: plasmid

Query Match 52.6%; Score 41; DB 2; Length 247;  
Best Local Similarity 46.7%; Pred. No. 16;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 GDLKTQIDQVESTAG 15  
| : : : : :  
Db 160 GDLKTKVDVICKVCG 174

RESULT 10

B71677

histidine kinase sensor protein (barA) RP229 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C>Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000

C:Accession: B71677

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark

Nature 396, 133-140, 1998

A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499

A:Accession: B71677

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-497 <AND>

A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14692.1; PID:g386

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: barA; RP229

Query Match 52.6%; Score 41; DB 2; Length 497;  
Best Local Similarity 53.3%; Pred. No. 35;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 DLKTDQIDQVESTAGS 16  
| : : : : :  
Db 254 NLKRAIEKVEKTS 268

RESULT 11

G72593

hypothetical protein APE1216 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C:Accession: G72593

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:99310339

A:Accession: G72593

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-533 <RAW>

A:Cross-references: DDBJ:AF000061; NID:g5104821; PIDN:BAA80205.1; PID:d1043991; PID:g

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1216

Query Match 52.6%; Score 41; DB 2; Length 533;  
Best Local Similarity 66.7%; Pred. No. 38;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 DLKTDQIDQVEST 13  
| : : : : :  
Db 388 DLNTRLDQVAST 399

## RESULT 12

H86843  
chorismate synthase (EC 4.6.1.4) [imported] - Lactococcus lactis subsp. lactis (strain 1  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: H86843  
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich  
Genome Res. in press, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium.  
A:Reference number: A86625  
A:Accession: H86843  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-388 <S70>  
A:Cross-references: GB:AF005176; NID:q12724773; PIDN:AAK05850.1; GSPDB:GN00146  
A:Experimental source: Strain 111403  
C:Genetics:  
A:Gene: aroC  
C:Superfamily: chorismate synthase  
C:Keywords: phosphorus-oxygen lyase

Query Match 51.3%; Score 40; DB 2; Length 388;

Best Local Similarity 53.3%; Pred. No. 40;

Matches 8; Conservative 3; Mismatches 3; Indels 4; Gaps 0;

QY 1 GDLKTOIDQVESTAG 15

Db 47 GRMKIESDQVEITSG 61

## RESULT 13

T18917  
hypothetical protein C04F12.10 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18917  
R:Lloyd, C.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19044  
A:Accession: T18917  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-442 <WIL>  
A:Cross-references: EMBL:Z81461; PIDN:CAR03839.1; GSPDB:GN00019; CESP:C04F12.10  
A:Experimental source: clone C04F12  
C:Genetics:  
A:Gene: CESP:C04F12.10  
A:Map position: 1  
A:Introns: 30/2; 77/3; 318/1; 340/3

Query Match 51.3%; Score 40; DB 2; Length 442;

Best Local Similarity 53.8%; Pred. No. 46;

Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 GDLKTOIDQVEST 13

Db 208 GDLKTKIEQLAAS 220

## RESULT 14

S44809  
F44B9.8 protein - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Sep-1999  
C:Accession: S44809  
R:Antonacci-Fulton, L.  
submitted to the EMBL Data Library, September 1993  
A:Description: Sequence of the C. elegans cosmid F44B9.  
A:Reference number: S44807  
A:Accession: S44809

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-447 <ANT>

A:Cross-references: EMBL:L23648; NID:g388585; PIDN:AAA28028.1; PID:g388588

C:Genetics:

A:Introns: 17/3; 53/1; 114/2; 375/1

C:Superfamily: phage T4 DNA polymerase accessory protein 44

C:Keywords: nucleus

Query Match 51.3%; Score 40; DB 2; Length 447;

Best Local Similarity 50.0%; Pred. No. 47;

Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 GDLKTOIDQVESTAGS 16

Db 252 GDMRTVINTLOSTAMS 267

## RESULT 15

T11560  
pol polyprotein - simian immunodeficiency virus SIIVsm (strain E543) (fragment)  
C:Species: simian immunodeficiency virus SIIVsm  
A:Variety: strain E543  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: T11560  
R:Hirsch, V.M.; Adger-Johnson, D.; Cambell, B.; Goldstein, S.; Brown, C.; Elkins, W.R.  
J. Virol. 71, 1608-1620, 1997  
A:Title: A molecularly cloned, pathogenic, neutralization-resistant simian immunodeficiency  
A:Reference number: Z17285; MUID:97151152  
A:Accession: T11560  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1019 <HIR>  
A:Cross-references: EMBL:U72748; NID:g1695908; PIDN:AAC56559.1; PID:g1695910  
C:Genetics:  
A:Gene: pol  
C:Superfamily: pol polyprotein  
C:Keywords: AIDS; immunodeficiency

Query Match 51.3%; Score 40; DB 2; Length 1019;

Best Local Similarity 57.1%; Pred. No. 1.2e+02;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 LKTOIDQVESTAGS 16

Db 884 LKTOIDRIEQANS 897

Search completed: July 5, 2001, 11:48:35

Job time: 374 sec



7

THIS PAGE BLANK (USPTO)



```
FT TRANSMEM      824      844      POTENTIAL.
FT DOMAIN        845      850      CYTOPLASMIC (POTENTIAL).
FT DOMAIN        811      816      CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT SEQUENCE      850 AA; 94538 MW; D653CAA39C93D2AC CRC64;
SQ SEQUENCE      850 AA; 94538 MW; D653CAA39C93D2AC CRC64;

Query Match      59.0%; Score 46; DB 1; Length 850;
Best Local Similarity 53.3%; Pred. No. 4.1;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVESTAG 15
    |||| :||| |
Db 471 GDLKARVDQVRQATG 485

RESULT 2
DEXT_STRSL
ID DEXT_STRSL STANDARD; PRT; 822 AA.
AC Q59979;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DEXTRANASE PRECURSOR (EC 3.2.1.11) (ALPHA-1,6-GLUCAN-6-
DE DEXTRANASE PRECURSOR (EC 3.2.1.11) (ALPHA-1,6-GLUCAN-6-
DE DEX OR DEXS.
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN SEQUENCE FROM N.A.
RP STRAIN=M-33;
RX MEDLINE=95255675; PubMed=7737522;
RA Ohnishi Y., Kubo S., Ono Y., Nozaki M., Gonda Y., Okano H.,
RA Matsuya T., Matsushiro A., Morita T.;
RT "Cloning and sequencing of the gene coding for dextranase from
RT Streptococcus salivarius.";
RL Gene 156:93-96(1995).
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,6-ALPHA-D-GLUCOSIDIC
CC LINKAGES IN DEXTRAN.
CC -!- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D29644; BAA06127.1; -.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 38 POTENTIAL.
FT CHAIN 39 822 DEXTRANASE.
SQ SEQUENCE 822 AA; 87924 MW; 31857D38D3F2C761 CRC64;

Query Match      53.8%; Score 42; DB 1; Length 822;
Best Local Similarity 53.3%; Pred. No. 19;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVESTAG 15
    |||| :||| |
Db 334 GDLKARIDMVNRKGTG 348

RESULT 3
GP6D_CHLTR
ID GP6D_CHLTR STANDARD; PRT; 247 AA.
AC P10560; P08787; Q46433;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
```

```
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE VIRULENCE PLASMID PROTEIN PGP6-D (PROTEIN P-10).
OS Chlamydia trachomatis.
OG Plasmid pLGV440, Plasmid pCHL1, and Plasmid pCTT1.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN SEQUENCE FROM N.A.
RP STRAIN=L2/434/BU; PLASMID=PLGV440;
RX MEDLINE=89013895; PubMed=2845228;
RA Comanducci M., Ricci S., Ratti G.;
RT "The structure of a plasmid of Chlamydia trachomatis believed to be
RT required for growth within mammalian cells.";
RL Mol. Microbiol. 2:531-538(1988).
RN SEQUENCE FROM N.A.
RP STRAIN=L1/440/LN; PLASMID=PLGV440;
RX MEDLINE=88233998; PubMed=2838808;
RA Hatt C., Ward M.E., Clarke I.N.;
RT "Analysis of the entire nucleotide sequence of the cryptic plasmid of
RT Chlamydia trachomatis serovar L1. Evidence for involvement in DNA
RT replication.";
RL Nucleic Acids Res. 16:4053-4067(1988).
RN SEQUENCE FROM N.A.
RP STRAIN=G0/86 / SEROTYPE D; PLASMID=pCHL1;
RX MEDLINE=90301796; PubMed=2194229;
RA Comanducci M., Ricci S., Cevenini R., Ratti G.;
RT "Diversity of the Chlamydia trachomatis common plasmid in biovars
RT with different pathogenicity.";
RL Plasmid 23:149-154(1990).
RN SEQUENCE FROM N.A.
RP STRAIN=SEROTYPE B; PLASMID=pCTT1;
RX MEDLINE=88177106; PubMed=344859;
RA Sriprakash K.S., Macavoy E.S.;
RT "Characterization and sequence of a plasmid from the trachoma biovar
RT of Chlamydia trachomatis.";
RL Plasmid 18:205-214(1987).
CC -!- MISCELLANEOUS: PGP6-D IS REQUIRED FOR GROWTH WITHIN MAMMALIAN
CC CELLS.
CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PLASMID PLGV440.
CC -!- SIMILARITY: BELONGS TO THE UPF0137 (PGP6-D) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07547; CAA30426.1; -.
KW Plasmid.
FT VARIANT 9 9 N -> D (IN PLASMIDS PCHL1 AND PCTT1).
FT VARIANT 47 47 Q -> R (IN PLASMIDS PCHL1 AND PCTT1).
FT VARIANT 183 186 VLDQ -> CWIE (IN PLASMID PCTT1).
FT VARIANT 194 194 K -> T (IN PLASMID PCTT1).
FT VARIANT 205 205 S -> L (IN PLASMID PCTT1).
FT VARIANT 222 222 M -> I (IN PLASMID PCTT1).
SQ SEQUENCE 247 AA; 28307 MW; F1FALIF29037C221 CRC64;

Query Match      52.6%; Score 41; DB 1; Length 247;
Best Local Similarity 46.7%; Pred. No. 7.6;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVESTAG 15
```

Db 160 GDLKTRVDVIGKVC 174  
|||||:|:|:|

RESULT 4  
AC11\_CAEEL STANDARD; PRT; 447 AA.  
AC P34429;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE PUTATIVE ACTIVATOR 1 36 KDA SUBUNIT (REPLICATION FACTOR C 36 KDA  
DE SUBUNIT) (A1 36 KDA SUBUNIT) (RF-C 36 KDA SUBUNIT) (RFC36).  
GN F4B9.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peleoderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Almsough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Watson J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
CC -1- FUNCTION: THE ELONGATION OF PRIMED DNA TEMPLATES BY DNA POLYMERASE  
CC DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS  
CC PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) AND ACTIVATOR 1 (BY  
CC SIMILARITY).  
CC -1- SUBUNIT: HETEROPTAMER OF SUBUNITS OF 140/145, 40, 38, 37, AND  
CC 36.5 KDA THAT FORMS A COMPLEX WITH PCNA IN THE PRESENCE OF ATP (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
CC -1- SIMILARITY: BELONGS TO THE ACTIVATOR 1 36 TO 40 KDA SUBUNITS  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: L23648; AAA28028.1;  
CC DR L23648; AAA28028.1;  
CC DR PIR: S44809; S44809.  
CC DR WormPep: F44B9.8; CE00554.  
KW Hypothetical protein; DNA replication; ATP-binding; Nuclear protein.  
FT NP\_BIND 72 79 ATP (POTENTIAL).  
SQ SEQUENCE 447 AA; 49825 MW; 8E5CE68DIAE131D2 CRC64;

Query Match 51.3%; Score 40; DB 1; Length 447;  
Best Local Similarity 50.0%; Pred. No. 21;  
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GDLKTRVDVIGKVC 174  
Db 252 GDMRTVINTLQSTAMS 267  
|||||:|:|:|  
RESULT 5

POL\_SIVS4  
ID POL\_SIVS4 STANDARD; PRT; 1019 AA.  
AC P12502;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Simian immunodeficiency virus (F236/smh4 isolate) (sooty mangabey).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11737;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89262053; PubMed=2786147;  
RA Hirsch V.M., Olmstead R.A., Murphy-Corb M., Purcell R.H.,  
RA Johnson P.R.;  
RT "An African primate lentivirus (SIVsm) closely related to HIV-2.";  
RL Nature 339:389-392(1989).  
CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: X14307; -; NOT\_ANNOTATED\_CDS.  
CC DR HSSP: P04584; IJLD.  
CC DR HIV: X14307; POLSMMH4.  
CC DR MEROPS: A02.002; -;  
CC DR InterPro: IPR000477; -;  
CC DR InterPro: IPR001037; -;  
CC DR InterPro: IPR001584; -;  
CC DR InterPro: IPR001969; -;  
CC DR InterPro: IPR001995; -;  
CC DR InterPro: IPR002156; -;  
CC DR Pfam: PF02022; Integrase\_Zn; 1.  
CC DR Pfam: PF00552; Integrase; 1.  
CC DR Pfam: PF00075; rnaaseh; 1.  
CC DR Pfam: PF00665; rve; 1.  
CC DR Pfam: PF00077; rvp; 1.  
CC DR Pfam: PF00078; rvt; 1.  
CC DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
CC DR PROSITE: PS00175; ASP\_PROT\_RETROV; 1.  
KW AIDS; Polyprotein; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase.  
FT CHAIN 1 167 PROTEASE.  
FT ACT\_SITE 93 93 BY SIMILARITY.  
SQ SEQUENCE 1019 AA; 115465 MW; 8D3DE0B85FC92B1C CRC64;

Query Match 51.3%; Score 40; DB 1; Length 1019;  
Best Local Similarity 57.1%; Pred. No. 52;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 3 LKTQIDQVESTAGS 16  
Db 884 LKTQIDRIRIQANS 897  
|||||:|:|:|  
RESULT 6  
POL\_SIVSP  
ID POL\_SIVSP STANDARD; PRT; 1022 AA.  
AC P19505; Q88140;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE POL POLYPROTEIN [CONTAINS: PROTEASE (RETROPEPSIN) (EC 3.4.23.16);  
DE REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].  
GN POL.  
OS Simian immunodeficiency virus (PBj14/BCL-3 isolate) (sooty mangabey).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11739;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90272009; PubMed=1971917;  
RA Dewhurst S., Embretson J.E., Anderson D.C., Mullins J.I., Fultz P.N.;  
RT "Sequence analysis and acute pathogenicity of molecularly cloned  
RT SIVSM-PBj14";  
RL Nature 345:636-640(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92368737; PubMed=1503826;  
RA Dewhurst S., Embretson J.E., Fultz P.N., Mullins J.I.;  
RT "Molecular clones from a non-acutely pathogenic derivative of  
RT SIVSM-PBj14: characterization and comparison to acutely pathogenic  
RT clones";  
RL AIDS Res. Hum. Retroviruses 8:1179-1187(1992).  
CC -1- P1M: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M31325; AAA47753.1; -;  
DR EMBL; L03298; AAA47777.1; -;  
DR HSP; P04584; IJLD.  
DR HIV; M31325; POLSSMMPBJ.  
DR MEROPS; A02.002; -;  
DR InterPro; IPR000477; -;  
DR InterPro; IPR001037; -;  
DR InterPro; IPR001584; -;  
DR InterPro; IPR001969; -;  
DR InterPro; IPR001995; -;  
DR InterPro; IPR002156; -;  
DR Pfam; PF02022; Integrase.Zn; 1.  
DR Pfam; PF00552; integrase; 1.  
DR Pfam; PF00075; rnaseh; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; rvp; 1.  
DR Pfam; PF00078; rvt; 1.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
KW AIDS; Polypeptide; Hydrolase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase.  
FT CHAIN 1 170  
FT ACT\_SITE 96 96 BY SIMILARITY.  
SQ SEQUENCE 1022 AA; 115869 MW; 16DFBEA03F289D6A CRC64;  
  
Query Match 51.3%; Score 40; DB 1; Length 1022;  
Best Local Similarity 57.1%; Pred. No. 52;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 3 LKTDQIDVESTAGS 16  
Db 887 LKTDQIDIREQANS 900  
|||||:|  
  
RESULT 7  
YMR9\_YEAST STANDARD; PRT; 523 AA.  
ID YMR9\_YEAST  
AC Q05040;

DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HYPOTHETICAL 59.3 KDA PROTEIN IN TAP42-CYK2 INTERGENIC REGION.  
GN YMR029C OR YMR973.02C.  
OS Saccharomycetes cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288C / AB972;  
RA Bowman S., Barrell B.G., Rajandream M.A.;  
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z49213; CAA89144.1; -;  
DR SGD; S0004631; YMR029C.  
KW Hypothetical protein; Coiled coil.  
FT DOMAIN 26 76 COILED COIL (POTENTIAL);  
SQ SEQUENCE 523 AA; 59301 MW; 3B2A3C95B2D575AE CRC64;  
  
Query Match 50.0%; Score 39; DB 1; Length 523;  
Best Local Similarity 53.3%; Pred. No. 37;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 DLKTDQIDVESTAGS 16  
Db 55 DLKHQLNQIQSKAVS 69  
|||||:|  
  
RESULT 8  
DNAK\_MYCHV STANDARD; PRT; 600 AA.  
ID DNAK\_MYCHV  
AC Q49539;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE DNAK PROTEIN (HEAT SHOCK 70 KDA PROTEIN) (65 KDA PROTEIN) (P65).  
GN DNAK.  
OS Mycoplasma hyopneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2099;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=232;  
RA Chou S.Y., Shiuan D.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
CC -1- INDUCTION: BY STRESS CONDITIONS E.G. HEAT SHOCK (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U50209; AAB01921.1; -;  
DR HSP; P04475; 1DKY.  
DR InterPro; IPR001023; -;  
DR Pfam; PF00012; HSP70; 1.

DR PRINTS: PR00301; HEATSHOCK70.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
KW Chaperone; ATP-binding; Heat shock.  
SQ SEQUENCE 600 AA; 65575 MW; 1F928F72EABDB6B CRC64;

Query Match 50.0%; Score 39; DB 1; Length 600;  
Best Local Similarity 46.7%; Pred. No. 43;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 DLKTDQIDQVESTAGS 16  
: : : : :  
Db 555 ELKLDQIEAAQS 569

## RESULT 9

PHBC\_RHIME STANDARD; PRT; 611 AA.  
AC P50176; O87321;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE POLY-BETA-HYDROXYBUTYRATE POLYMERASE (EC 2.3.1.-) (POLY(3-HYDROXYBUTYRATE) POLYMERASE) (PHB POLYMERASE) (PHB SYNTHASE)  
DE DEHYDROXYALKANOATE POLYMERASE) (PHA POLYMERASE) (PHA SYNTHASE)  
-GN (POLYHYDROXYALKANOIC ACID SYNTHASE).  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE OF 35-611 FROM N.A.  
RC STRAIN=41;  
RX MEDLINE=96036213; PubMed=7582015;  
RA Tomboloni R., Povolito S., Buson A., Squartini A., Nuti M.P.;  
RT "Poly-beta-hydroxybutyrate (PHB) biosynthetic genes in Rhizobium meliloti 41."  
RL Microbiology 141:2553-2559 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=98405222; PubMed=9734305;  
RA Willis L.B., Walker G.C.;  
RT "The phbC (poly-beta-hydroxybutyrate synthase) gene of Rhizobium (Sinorhizobium) meliloti and characterization of phbC mutants.";  
RL Can. J. Microbiol. 44:554-564 (1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RA Galibert F., Capela D., Hubler-Barloy F., Gatiou M., Batut J., Boiscard P., Gouzy J., Kahn D., Thebaud P., Goffeau A., Purnelle B., Pohl T., Bothe G., Schneider S., Portetelie D., Vandenbol M., Puehler A., Becker A., Weidner S.;  
RL Submitted (MAR-2000) to the SWISS-PROT data bank.  
CC -1- FUNCTION: POLYMERIZES D(-)-3-HYDROXYBUTYRYL-COA TO CREATE PHB WHICH CONSISTS OF THOUSANDS OF HYDROXYBUTYRATE MOLECULES LINKED WHEN CELLS GROW UNDER CONDITIONS OF NUTRIENT LIMITATION.  
CC WHEN CELLS GROW UNDER CONDITIONS OF NUTRIENT LIMITATION.  
CC -1- SUBUNIT: MONOMER (PROBABLE).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO THE PHA/PHB SYNTHASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

CC EMBL; U17227; AAA90984.1; -  
DR EMBL; AF031938; AAC61899.1; -  
DR InterPro; IPR000073; -  
DR Pfam; PF00561; abhydrolase; 1.  
KW PHB biosynthesis; Transferase; Acyltransferase.  
FT ACT\_SITE 349 349  
FT CONFLICT 111 111 G -> A (IN REF. 1).  
FT CONFLICT 122 122 R -> H (IN REF. 1).  
FT CONFLICT 157 157 R -> K (IN REF. 1).  
FT CONFLICT 477 477 K -> R (IN REF. 1).  
FT CONFLICT 481 481 V -> M (IN REF. 1).  
SQ SEQUENCE 611 AA; 68039 MW; ABE8A9B17F87D7A9 CRC64;

Query Match 50.0%; Score 39; DB 1; Length 611;  
Best Local Similarity 43.8%; Pred. No. 44;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps

OY 1 GDLKTDQIDQVESTAGS 16  
: : : : :  
Db 555 GDIETWGMKAKETAGS 570

## RESULT 10

Y136\_HUMAN STANDARD; PRT; 950 AA.  
AC Q14149; Q9UEZ2;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE HYPOTHETICAL PROTEIN KIAA0136 (FRAGMENT).  
GN KIAA0136.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Bone marrow;  
RX MEDLINE=96127530; PubMed=8590280;  
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. IV. The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by analysis of cDNA clones from human cell line KG-1."  
RL DNA Res. 2:167-174 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Shibuya K., Kudo J., Minoshima S., Kawasaki K., Nakatoh E., Shintani A., Asakawa S., Shimizu N.;  
RT "Genomic sequencing of 1.2-Mb region on human chromosome 21q22.2.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D50926; BAA09485.1; -  
DR EMBL; AP000693; BAA89432.1; -  
DR EMBL; AP000692; BAA89432.1; JOINED.  
KW Hypothetical protein.  
FT CONFLICT 100 100 L -> V (IN REF. 2).  
FT NON\_TER 1  
SQ SEQUENCE 950 AA; 108302 MW; 70F9E806655ACC2 CRC64;  
Query Match 50.0%; Score 39; DB 1; Length 950;  
Best Local Similarity 50.0%; Pred. No. 71;  
Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 DLKTDQVEST 13
DB 864 ELKTEVEQLKST 875

RESULT 11
BAG_STRAG STANDARD; PRT: 1164 AA.
AC P27951;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IGA FC RECEPTOR PRECURSOR (BETA ANTIGEN) (B ANTIGEN).
GN BAG.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.
RP STRAIN=LA239;
RX MEDLINE=91312121; PubMed=1857207;
RA Jerlstrom P.G., Chhatwal G.S., Timmis K.N.;
RT "The IgA-binding beta antigen of the c protein complex of Group B
RT streptococci: sequence determination of its gene and detection of two
RT binding regions."
RL Mol. Microbiol. 5:843-849(1991).
RN [2]
RP IDENTIFICATION OF IG-LIKE DOMAIN.
RX MEDLINE=97035265; PubMed=8880921;
RA Bateman A., Eddy S.R., Chothia C.;
RT "Members of the immunoglobulin superfamily in bacteria."
RL Protein Sci. 5:1939-1942(1996).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59771; CAA42442.1;
CC PIR: S15330; FCSOAG.
CC InterPro: IPR001899;
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC Cell wall; Transmembrane; Receptor; Repeat; Signal;
CC Immunoglobulin domain.
FT SIGNAL 1
FT CHAIN 37
FT TRANSMEM 38 1164 IGA FC RECEPTOR.
FT DOMAIN 38 1131 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1141 1159 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 1160 1164 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 434 534 IG-LIKE DOMAIN.
FT DOMAIN 199 438 IGA-BINDING (POTENTIAL).
FT TRANSMEM 439 826 IGA-BINDING (POTENTIAL).
FT DOMAIN 827 945 PRO-RICH REPEATS.
FT TRANSMEM 1131 1137 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT DOMAIN 1131 1137 PROTEINS.
FT SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;

Query Match 50.08; Score 39; DB 1; Length 1164;
Best Local Similarity 50.08; Pred. No. 89;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 DLKTDQVESTAG 15
DB 84 DIKRAVEPVENTAG 97

RESULT 12
T2D7_DROME
ID T2D7_DROME STANDARD; PRT: 278 AA.
AC Q27272; O9VX16;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSCRIPTION INITIATION FACTOR TFIID 42 KDA SUBUNIT (TAFII-42)
DE (TAFII40) (P42) (ENHANCER OF YELLOW 1 PROTEIN).
GN E(Y)1 OR TAF40 OR CG6474.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=94150630; PubMed=7545910;
RX Kokubo T., Gong D.W., Wootton J.C., Horikoshi M., Roeder R.G.,
RA Nakatani Y.;
RT "Molecular cloning of Drosophila TFIID subunits."
RL Nature 367:484-487(1994).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=OREGON-R;
RX MEDLINE=94037099; PubMed=8221891;
RA Goodrich J.A., Hoey T., Thut C.J., Admon A., Tjian R.;
RT "Drosophila TAFII40 interacts with both a VP16 activation domain and
RT the basal transcription factor TFIIB."
RL Cell 75:519-530(1993).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arif J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shier B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```



RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
CC -1- FUNCTION: TAFs ARE COMPONENTS OF THE TRANSCRIPTION FACTOR IID  
CC (TFIID) COMPLEX THAT ARE ESSENTIAL FOR MEDIATING REGULATION OF  
CC RNA POLYMERASE TRANSCRIPTION (BY SIMILARITY).  
CC -1- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A  
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: BELONGS TO THE TAF2G FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U06458; AAC47347.1; -  
CC EMBL; L29540; AAA28488.1; -  
CC EMBL; AE003506; AAF48767.1; -  
CC HSP; P30129; 4DPV.  
CC TRANSFAC; T02125; -  
CC FlyBase; FBgn0000617; e(y)1.  
CC Transcription regulation; Nuclear protein.  
CC DOMAIN 190 197 POLY-THR.  
CC DOMAIN 202 209 POLY-GLY.  
CC DOMAIN 230 236 POLY-GLY.  
CC DOMAIN 267 274 POLY-GLU (ACIDIC).  
CC SEQUENCE 278 AA; 29314 MW; 0EA442C80467001F CRC64;  
SQ  
Query Match 48.7%; Score 38; DB 1; Length 278;  
Best Local Similarity 43.8%; Pred. No. 28;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 1 GDLKTDQDVSTAGS 16  
|||||:|:|:|:  
DB 221 GDLKMEVSDAAVGS 236  
RESULT 13  
SAHL RHOSH STANDARD; PRT; 463 AA.  
AC O50562;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE  
DE HYDROLASE) (ADHCYASE).  
GN AHCY.  
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;  
OC Rhodobacter.  
OX NCBI\_TaxID=1063;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97354111; PubMed=9210332;  
RA Mizoguchi H., Masuda T., Nishimura K., Shimada H., Ohta H., Shioi Y.,  
RA Takamiya K.;  
RT "Nucleotide sequence and transcriptional analysis of the flanking  
RT region of the gene (spb) for the trans-acting factor that controls  
RT light-mediated expression of the puf operon in Rhodobacter  
RT sphaeroides";  
RL Plant Cell Physiol. 38:558-567(1997).  
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =  
CC ADENOSINE + L-HOMOCYSTEINE.  
CC -1- COFACTOR: NAD (BY SIMILARITY).  
CC -1- PATHWAY: ACTIVATED METHYL CYCLE.  
CC -1- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U76671; AAB88245.1; -  
CC InterPro; IPR000043; -  
CC Pfam; PF00670; AdoHcyase\_1;  
CC PROSITE; PS00738; AdoHcyase\_1; 1.  
CC PROSITE; PS00739; AdoHcyase\_2; 1.  
CC Hydrolase; NAD; One-carbon metabolism.  
FT NP\_BIND 247 278 NAD (POTENTIAL).  
SQ SEQUENCE 463 AA; 50617 MW; ACE8EB9D0DCCBBA CRC64;  
Query Match 48.7%; Score 38; DB 1; Length 463;  
Best Local Similarity 40.0%; Pred. No. 48;  
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps  
QY 2 DLKTDQDVSTAGS 16  
:|||:|:|:|:  
DB 351 NIKDQDVIEMPSGS 365  
RESULT 14  
YG3A\_YEAST STANDARD; PRT; 816 AA.  
AC P53278;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE HYPOTHETICAL 92.7 KDA PROTEIN IN ASN2-PHBI INTERGENIC REGION.  
GN YGR130C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA van Dyck L., Skala J., de Wergifosse P., Purnelle B., Tallia E.,  
RA Nawrocki A., del Bino S., Goffeau A.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaborat  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; 272915; CAA97143.1; -  
CC HSSP; P56255; 1PJR.  
CC SGD; S0003362; YGR130C.  
KW Hypothetical protein.  
FT DOMAIN 164 185 POLY-SER.  
FT DOMAIN 205 218 POLY-THR.  
SQ SEQUENCE 816 AA; 92698 MW; D482327C9274C026 CRC64;  
Query Match 48.7%; Score 38; DB 1; Length 816;  
Best Local Similarity 50.0%; Pred. No. 89;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
QY 2 DLKTDQDVST 13  
||:|:|:|:  
DB 460 DLQNQIDEIENS 471  
RESULT 15  
TEGU\_HSV6G

TEGU\_HSV6G STANDARD; PRT: 2077 AA.  
AC P30002;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCF-2000 (Rel. 40, Last annotation update)  
DE LARGE TEGUMENT PROTEIN.  
GN U31.  
OS Human herpesvirus (type 6 / strain GS) (HHV6).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Roseolovirus.  
OX NCBI\_TaxID=10369;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91374623; PubMed=1654455;  
RA Josephs S.F., Ablashi D.V., Salahuddin S.Z., Jagodzinski L.L.,  
RA Wong-Staal F., Gallo R.C.;  
RT "Identification of the human herpesvirus 6 glycoprotein H and  
RT putative large tegument protein genes.";  
RL J. Virol. 65:5597-5604(1991).  
CC -!- FUNCTION: TEGUMENT PROTEIN.  
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
CC EBV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: S57540; AAB19786.1; -.  
DR PIR: H40511; WZBE24.  
SQ SEQUENCE 2077 AA; 239909 MW; 28E297FDC2FAD2AE CRC64;

Query Match 48.7%; Score 38; DB 1; Length 2077;  
Best Local Similarity 50.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
Oy 2 DLKTOIDQVEST 13  
|:|:|:|:|:|:  
Db 1084 DVKTEIDQMOKS 1095

Search completed: July 5, 2001, 11:51:44  
Job time: 503 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:50:49 ; Search time 123.78 Seconds  
(without alignments)  
17.102 Million cell updates/sec

Title: US-09-462-480-9  
Perfect score: 78  
Sequence: 1 GDLTKIQDVSTAGS 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_16.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_prodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	100	2 069739	Q96739 mycobacteri
2	46	59.0	1163	10 Q9SHT1	Q9SHT1 arabidopsis
3	44	56.4	2954	14 Q96898	Q96898 hepatitis g
4	43.5	55.8	630	3 Q13870	Q13870 schizosacch
5	43	55.1	1088	2 Q9JW98	Q9JW98 neisseria m
6	43	55.1	1082	2 Q9JXM7	Q9JXM7 neisseria m
7	43	55.1	1084	2 Q9JPL3	Q9JPL3 neisseria m
8	42	53.8	600	11 Q9EPM6	Q9EPM6 mus musculu
9	42	53.8	2756	2 Q33904	Q33904 shevanella
10	41	52.6	437	2 Q9ZDU5	Q9ZDU5 rickettsia
11	41	52.6	533	1 Q9YCP2	Q9YCP2 aeropyrum p
12	41	52.6	675	5 Q9VYB1	Q9VYB1 drosophila
13	41	52.6	779	5 Q9W0U8	Q9W0U8 drosophila
14	41	52.6	1616	5 Q9V5J0	Q9V5J0 drosophila
15	40	51.3	271	2 Q32408	Q32408 rhodospiril
16	40	51.3	280	5 Q9VL68	Q9VL68 drosophila
17	40	51.3	442	5 Q9XVE5	Q9XVE5 caenorhabdi
18	40	51.3	533	9 Q80226	Q80226 bacterioph
19	40	51.3	1019	14 P89154	P89154 chimpanzee

20	40	51.3	1022	14	Q87956	Q87956 chimpanzee
21	40	51.3	1022	14	Q87965	Q87965 chimpanzee
22	40	51.3	1022	14	Q88135	Q88135 chimpanzee
23	40	51.3	1022	14	Q89620	Q89620 chimpanzee
24	40	51.3	1022	14	Q90317	Q90317 chimpanzee
25	40	51.3	1096	5	Q22489	Q22489 caenorhabdi
26	39	50.0	197	4	Q75596	Q75596 homo sapien
27	39	50.0	774	5	Q23424	Q23424 caenorhabdi
28	39	50.0	1134	2	Q99051	Q99051 streptococ
29	39	50.0	1390	10	Q49145	Q49145 arabidopsis
30	38	48.7	128	1	Q29884	Q29884 archaeoglob
31	38	48.7	197	6	Q28008	Q28008 bos taurus
32	38	48.7	265	2	Q9F9K3	Q9F9K3 borrelia bi
33	38	48.7	265	2	Q9F9K1	Q9F9K1 borrelia bi
34	38	48.7	334	2	Q68119	Q68119 rhodobacter
35	38	48.7	414	5	Q19088	Q19088 caenorhabdi
36	38	48.7	446	5	Q9W0C6	Q9W0C6 drosophila
37	38	48.7	550	2	Q69874	Q69874 streptomyce
38	38	48.7	550	2	Q85676	Q85676 streptomyce
39	38	48.7	555	10	Q9FI58	Q9FI58 arabidops.
40	38	48.7	585	8	Q19929	Q19929 cyanidium c
41	38	48.7	767	2	Q51268	Q51268 borrelia bu
42	38	48.7	1733	2	Q53579	Q53579 mycobacteri
43	38	48.7	1832	2	Q9L8C8	Q9L8C8 sorangium c
44	38	48.7	1832	2	Q9KIZ8	Q9KIZ8 sorangium c
45	38	48.7	2077	14	Q9WT33	Q9WT33 human herpes

## ALIGNMENTS

RESULT 1					
Q69739					
ID	Q69739	PRELIMINARY;	PRT;	100	AA.
AC	Q69739;				
DT	01-AUG-1998 (TREMBLrel. 07, Created)				
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)				
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)				
DE	HYPOTHETICAL 10.8 KDA PROTEIN.				
GN	LHP OR RV3874 OR MTV027.09.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,				
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RL	complete genome sequence."				
RN	Nature 393:537-544(1998).				
RP	[2]				
RC	SEQUENCE FROM N.A.				
RA	STRAIN=H37RV;				
RT	Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;				
RT	"Promoter analysis of the M. tuberculosis orfIC gene encoding the				
RL	early secreted antigenic target 6 kDa (ESAT-6).";				
DR	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF022120; CAA17966.1;				
DR	EMBL; AF004671; AAC83445.1;				
DR	TubercuList; Rv3874;				
KW	Hypothetical protein.				
SQ	SEQUENCE 100 AA; 10794 MW; 285F4FC96F55D194 CRC64;				

Query Match 100.0%; Score 78; DB 2; Length 100;  
Best Local Similarity 100.0%; Pred. No. 6.9e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVESTAGS 16

Db 23 GDLKTQIDQVESTAGS 38

RESULT 2

Q9SHT1 PRELIMINARY; PRT; 1163 AA.

AC Q9SHT1; -  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE PUTATIVE CHROMOSOME ASSOCIATED PROTEIN.  
GN AT2G2170;  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=20083487; PubMed=10617194;  
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
RA Buell C.R., Ketchum K.A., Lee J.J., Roming C.M., Koo H., Moffat K.S.,  
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,  
RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
RA Copenhaver G.P., Pruss D., Nierman W.C., White O., Eisen J.A.,  
RA Salzberg S.L., Fraser C.M., Venter J.C.;  
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";  
RL Nature 402:761-768(1999).  
DR EMBL; AC007290; AAD26882.2; -  
DR InterPro; IPR003395; -  
DR InterPro; IPR003405; -  
DR InterPro; IPR003439; -  
DR Pfam; PF02463; SMC\_N; 1.  
DR Pfam; PF02483; SMC\_C; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; UNKNOWN\_1.  
SQ SEQUENCE 1163 AA; 134366 MW; 775AD20AEA526B40 CRC64;

Query Match 59.0%; Score 46; DB 10; Length 1163;

Best Local Similarity 61.5%; Pred. No. 26;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVEST 13

Db 710 GDITRIDQVRSS 722

RESULT 3

Q96898 PRELIMINARY; PRT; 2954 AA.

AC Q96898;  
DT 01-FEB-1997 (TREMBlrel. 02, Created)  
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE GENOME POLYPROTEIN.  
OS Hepatitis GB virus A.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.  
OX NCBI\_TaxID=39112;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95241511; PubMed=7724574;  
RA Simons J.N., Pilot-Matias T.J., Leary T.P., Dawson G.J., Desai S.M.,  
RA Schlauder G.G., Muerhoff A.S., Erker J.C., Buijk S.L., Chalmers M.L.,  
RA van Sant C.L., Mushahwar I.K.;  
RT "Identification of two flavivirus-like genomes in the GB hepatitis

agent.";  
Proc. Natl. Acad. Sci. U.S.A. 92:3401-3405(1995).  
RN [2]  
RP SEQUENCE FROM N.A.

RX MEDLINE=96323131; PubMed=8709237;  
RA Simons J.N., Desai S.M., Schultz D.E., Lemon S.M., Mushahwar I.K.;  
RT "Translation initiation in GB viruses A and C: evidence for internal  
RT ribosome entry and implications for genome organization.";  
RL J. Virol. 70:6126-6135(1996).  
CC -1- SIMILARITY: TO HEPATITIS C VIRUS RNA DEPENDENT RNA POLYMERASE.  
DR EMBL; U22303; AAC55983.1; -  
DR HSSP; P27958; IHEI.  
DR MEROPS; S29.002; -  
DR InterPro; IPR000255; -  
DR InterPro; IPR000745; -  
DR InterPro; IPR001410; -  
DR InterPro; IPR001490; -  
DR InterPro; IPR002166; -  
DR InterPro; IPR002518; -  
DR InterPro; IPR002868; -  
DR Pfam; PF00998; HCV\_RdRP; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN\_1.  
DR SMART; SM00487; DEXdc; 1.  
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase.

SQ SEQUENCE 2954 AA; 318835 MW; 4A0C35FC2ED283B1 CRC64;

Query Match 56.4%; Score 44; DB 14; Length 2954;

Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVESTAGS 16

Db 2379 GDITRIDQIEETPTS 2394

RESULT 4

Q13870 PRELIMINARY; PRT; 630 AA.

AC Q13870;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-JUN-1998 (TREMBlrel. 06, Last annotation update)  
DE PROBABLE TRANSCRIPTIONAL REGULATOR ClB3.05.  
GN SPAC1B3.05.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -1- SIMILARITY: TO YEAST NOT3.  
DR EMBL; Z98598; CAB11234.1; -  
KW Nuclear protein; Transcription regulation.  
FT DOMAIN 235 243 POLY-SER.  
FT DOMAIN 243 252 POLY-SER.  
FT DOMAIN 374 385 POLY-ALA.  
FT DOMAIN 624 627 POLY-ASP.  
SQ SEQUENCE 630 AA; 4F9284439845CB71 CRC64;

Query Match 55.8%; Score 43.5; DB 3; Length 630;

Best Local Similarity 47.8%; Pred. No. 36;  
Matches 11; Conservative 3; Mismatches 2; Indels 7; Gaps 1;

```
QY 1 GDLKTOI-----DOVESTAGS 16
RL |||||
DB 35 GDLKTIKKLQRLRDQIKTWASS 57

RESULT 5
Q9JW98 PRELIMINARY; PRT; 1068 AA.
AC Q9JW98;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PEPTIDASE (EC 3.4.21.).
GN NMA0478.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL; AL162753; CAB83773.1;
DR InterPro; IPR000209;
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 1068 AA; 111854 MW; D8213CEAB86206A1 CRC64;

Query Match 55.1%; Score 43; DB 2; Length 1068;
Best Local Similarity 56.2%; Pred. No. 76;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GDLKTIQIDQVESTAGS 16
RL |||||
DB 654 GGLLASLDSVEKTAGS 669

RESULT 6
Q9JXM7 PRELIMINARY; PRT; 1082 AA.
AC Q9JXM7;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE SEROTYPE-1-SPECIFIC ANTIGEN, PUTATIVE.
GN NMB1969.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
```

```
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002345; AAF42298.1;
DR TIGR; NMB1969;
DR InterPro; IPR000209;
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
SQ SEQUENCE 1082 AA; 113614 MW; 3779C7177FAF056 CRC64;

Query Match 55.1%; Score 43; DB 2; Length 1082;
Best Local Similarity 56.2%; Pred. No. 78;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GDLKTIQIDQVESTAGS 16
RL |||||
DB 668 GGLLASLDSVEKTAGS 683

RESULT 7
Q9JPL3 PRELIMINARY; PRT; 1084 AA.
AC Q9JPL3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE AUTOTRANSPORTER SERINE PROTEASE.
GN AUP.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B:15:PI.16;
RA Turner D., Wooldridge K.G., Ala'Aldeen D.A.A.;
RT "Identification and characterisation of an autotransporter serine
RT protease in Neisseria meningitidis."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277537; CAB89119.1;
DR InterPro; IPR000209;
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 1084 AA; 113743 MW; 33BC30E9543D0AC3 CRC64;
```

```
Query Match 55.1%; Score 43; DB 2; Length 1084;
Best Local Similarity 56.2%; Pred. No. 78;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GDLKTIQIDQVESTAGS 16
RL |||||
DB 670 GGLLASLDSVEKTAGS 685

RESULT 8
Q9EPM6 PRELIMINARY; PRT; 600 AA.
AC Q9EPM6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE FYVE-FINGER CONTAINING PROTEIN (FRAGMENT).
GN RABIP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

## RC TISSUE-SKELETAL MUSCLE;

RA Cormont M., Mari M., Galmiche A., Le Marchand-Brustel Y.,  
 RT "Rabip4, a new FYVE-finger containing protein, is a Rab4 effector.  
 RL Involved in the fusion between Rab4 and Rab5 positive endosomes."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ250024; CAC17732.1; -  
 FT NON\_TER 600  
 SQ SEQUENCE 600 AA; 69054 MW; A2580582CC1249B4 CRC64;

Query Match 53.8%; Score 42; DB 11; Length 600;  
 Best Local Similarity 61.5%; Pred. No. 61;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVEST 13

Db 215 GDLQKIDGLEXT 227

## RESULT 9

ID O33904 PRELIMINARY; PRT; 2756 AA.  
 AC O33904;

DT 01-JAN-1998 (TRENBLrel. 05, Created)

DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE HYPOTHETICAL 292.7 KDA PROTEIN.

OS Shewanella sp. SCRC-2738.

OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;

CC Shewanella.

OX NCBI\_TaxID=53560;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SCRC-2738;

EX MEDLINE=97419510; PubMed=9274025;

RA Takeyama H., Takeda K., Yazawa K., Yamada A., Matsunaga T.;

RT "Expression of the eicosapentaenoic acid synthesis gene cluster from  
 RT Shewanella sp. in a transgenic marine cyanobacterium, Synechococcus  
 sp.";

RL Microbiology 143:0-0(0).

DR EMBL; U73935; AAB81123.1; -

DR InterPro; IPR000255; -

DR InterPro; IPR000794; -

DR InterPro; IPR001227; -

DR Pfam; PF00109; ketoacyl-synt; 1.

DR Pfam; PF00550; pp-binding; 6.

DR Pfam; PF00698; Acyl trans; 1.

DR PROSITE; PS50075; ACP\_DOMAIN; 6.

KW Hypothetical protein; Phosphopantetheine.

SQ SEQUENCE 2756 AA; 292671 MW; 910FF442DB5F7759 CRC64;

## Query Match

53.8%; Score 42; DB 2; Length 2756;

Best Local Similarity 50.08; Pred. No. 3.le+02;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GDLKTQIDQVESTAGS 16

Db 397 GSVKSGIKTKSTAGT 412

## RESULT 10

ID Q9ZDU5

AC Q9ZDU5;

PRELIMINARY; PRT; 497 AA.

DT 01-MAY-1999 (TRENBLrel. 10, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE HISTIDINE KINASE SENSOR PROTEIN (BARA).

GN RP229.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

CC Rickettsiaceae; Rickettsia.

OX NCBI\_TaxID=782;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MADRID E;

RX MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,

RA Sichertitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of

RT mitochondria";

RL Nature 396:133-140(1998).

CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE

CC KINASES.

DR EMBL; AJ235271; CAA14692.1; -

DR InterPro; IPR000410; -

DR InterPro; IPR003594; -

DR Pfam; PF00512; signal; 1.

DR PRINTS; PR00344; BCTRLSENSOR.

DR SMART; SM00387; HATPase\_C; 1.

KW Kinase; Phosphorylation; Sensory transduction; Transferase.

SQ SEQUENCE 497 AA; 56908 MW; 781C9147531F388D CRC64;

## Query Match

52.6%; Score 41; DB 2; Length 497;

Best Local Similarity 53.3%; Pred. No. 74;

Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 DLKTQIDQVESTAGS 16

Db 254 NLKAQIEKVEKTSST 268

## RESULT 11

QYICP2

ID QYICP2 PRELIMINARY; PRT; 533 AA.

AC QYICP2;

DT 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE HYPOTHETICAL 56.5 KDA PROTEIN APE1216.

GN APE1216.

OS Aeropyrum pernix.

OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;

OC Aeropyrum.

OX NCBI\_TaxID=56636;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K1;

RX MEDLINE=99310339; PubMed=10382966;

RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,

RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,

RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,

RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,

RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,

RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;

RT "Complete genome sequence of an aerobic hyper-thermophilic

RT Crenarchaeon, Aeropyrum pernix K1.";

RL DNA Res. 6:83-101(1999).

DR EMBL; AP000061; BAA80205.1; -

DR InterPro; IPR000533; -

DR InterPro; IPR001865; -

DR InterPro; IPR002017; -

DR PRINTS; PR00194; TROPOMYOSIN.

KW PROSITE; PS00962; RIBOSOMAL\_S2\_1; UNKNOWN\_1.

SQ SEQUENCE 533 AA; 56481 MW; E8B53D791C353E31 CRC64;

## Query Match

52.6%; Score 41; DB 1; Length 533;

Best Local Similarity 66.7%; Pred. No. 80;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 DLKTQIDQVEST 13

```

Db 388 DLNTRLDQVAST 399
|||||
RESULT 12
Q9VYB1 PRELIMINARY; PRT; 675 AA.
ID Q9VYB1
AC Q9VYB1
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-MAY-2000 (TREMREL. 13, Last annotation update)
DE CG9938 PROTEIN.
GN CG9938
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Baller R.M., Basu A., Baxendale J., Andrews-Pfannkuch C., Baldwin D.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foslter C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harlis N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Kallush F., Kravitz S., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalousty A., Krapf C., Kravitz S., Kulp D., Lai Z.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merklov G., Milshina N.V., Mobarly C., McLeod M.P., McPherson D.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Pittman K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; AF003467; AAF47335.1;
DR EMBL; AF003467; AAF47336.1;
DR HSP; P29354; 1GRI.
DR FlyBase; FBgn0035101; p130CAS.
DR InterPro; IPR001452;
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PS00002; SH3; 1.
DR SMART; SM00326; SH3; 1.
KW Alternative splicing; Hypothetical protein.
FT VARSPPLIC 1 211 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 779 AA; 84174 MW; 68B883F198CA45F6 CRC64;

```

```

QY 1 GDLTKTDQVES 12
Db 309 GDLTKTDQVES 320
Query Match 52.6%; Score 41; DB 5; Length 675;
Best Local Similarity 58.3%; Pred. No. 1e+02;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

Query Match 52.6%; Score 41; DB 5; Length 779;  
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 DLKTOIDQVESTAGS 16  
 ||||| | | | |  
 Db 594 DLKTOFDVTIRTAES 608

## RESULT 14

Q9VSJ0 Q9VSJ0 PRELIMINARY; PRT; 1616 AA.  
 AC Q9VSJ0; Q9VSJ9;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DE ECDYSONE-INDUCIBLE GENE EL.  
 GN IMPEI OR CG7116 OR CG13668.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RN Science 287:2185-2195(2000).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO, AND IMAGINAL DISKS;  
 RA Clark K.E., Vesenska G.D., Robertson J.P., Natzke J.E.;  
 RT "Tissue specific regulation of the ecdysone-inducible gene IMP-EL";  
 CC submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A  
 (LDLRA) DOMAIN.

Query Match 52.6%; Score 41; DB 5; Length 1616;  
 Best Local Similarity 53.8%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 DLKTOIDQVESTA 14  
 |:::||||| |  
 Db 651 DMEIDQVEQSA 663

## RESULT 15

O32408 O32408 PRELIMINARY; PRT; 271 AA.  
 ID AC O32408;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DE 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE BACTERIOCHLOROPHYLL BIOSYNTHETIC ENZYME (FRAGMENT).  
 GN BCHZ.  
 OS Rhodospirillum rubrum.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;  
 OC Phaeospirillum.  
 OX NCBI\_TaxID=1083;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nagashima K.V., Matsuura K., Shimada K.;  
 RL Photosyn. Res. 50:61-70(1996).  
 DR EMBL; D50654; BAA22794.1; -  
 FT NON\_TER 1  
 SQ SEQUENCE 271 AA; 29711 MW; D483D34750FB2DF4 CRC64;

Query Match 51.3%; Score 40; DB 2; Length 271;  
 Best Local Similarity 72.7%; Pred. No. 57;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 LKTOIDQVEST 13  
 |:::||||| |  
 Db 188 LGTQDQVEAT 198

Search completed: July 5, 2001, 11:50:50  
 Job time: 473 sec



7

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:45:43 ; Search time 130.35 seconds  
(without alignments)  
9.767 Million cell updates/sec

Title: US-09-462-480-8  
Perfect score: 107  
Sequence: 1 QEAGNFERISGLKVTQIDQV 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues  
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89.5	83.6	20	AAV03708	M. tuberculosis LH
2	89.5	83.6	49	AAV03706	M. tuberculosis LH
3	89.5	83.6	95	AAW32444	Mycobacterium tube
4	89.5	83.6	95	AAW32376	Mycobacterium tube
5	89.5	83.6	95	AAW81747	M. tuberculosis im
6	89.5	83.6	95	AAW64321	Mycobacterium tube
7	89.5	83.6	95	AAV32097	Mycobacterium tube
8	89.5	83.6	95	AAV39118	M. tuberculosis an
9	89.5	83.6	95	AAV38981	M. tuberculosis re
10	89.5	83.6	100	AAW81706	M. tuberculosis im
11	89.5	83.6	100	AAW64339	Mycobacterium tube

12	89.5	83.6	100	AAV39136	M. tuberculosis an
13	89.5	83.6	100	AAV38993	M. tuberculosis re
14	89.5	83.6	100	AAV03705	M. tuberculosis LH
15	89.5	83.6	100	AAW35218	M. tuberculosis RV3
16	89.5	83.6	100	AAW19845	Mycobacterium tube
17	89.5	83.6	802	AAW81746	M. tuberculosis fu
18	89.5	83.6	802	AAW64379	Mycobacterium anti
19	89.5	83.6	802	AAV32063	Mycobacterium tube
20	89.5	83.6	802	AAV39224	M. tuberculosis fu
21	89.5	83.6	802	AAV39176	M. tuberculosis fu
22	89.5	83.6	802	AAV39081	M. tuberculosis fu
23	89.5	83.6	802	AAV39033	M. tuberculosis fu
24	74.5	69.6	28	AAW32455	Mycobacterium tube
25	74.5	69.6	28	AAW81693	M. tuberculosis im
26	74.5	69.6	28	AAV39123	M. tuberculosis an
27	47.5	44.4	80	AAW32454	Mycobacterium tube
28	47.5	44.4	80	AAW32386	Mycobacterium tube
29	47.5	44.4	80	AAW81707	M. tuberculosis im
30	47.5	44.4	80	AAW64340	Mycobacterium tube
31	47.5	44.4	80	AAV39137	M. tuberculosis an
32	47.5	44.4	80	AAV38994	M. tuberculosis re
33	46.5	43.5	360	AAW94375	Laminin A chain (a
34	46.5	43.5	360	AAW39205	Drosophila partial
35	43	40.2	227	AAV03765	S. aureus spoJ1 O
36	43	40.2	281	AAV03764	S. aureus spoJ1 P
37	43	40.2	671	AAW24591	Arabidopsis thalia
38	43	40.2	671	AAV39770	Arabidopsis thalia
39	43	40.2	673	AAW24590	Arabidopsis thalia
40	43	40.2	673	AAV39769	Arabidopsis thalia
41	43	40.2	682	AAW24589	Arabidopsis thalia
42	43	40.2	682	AAV39768	Arabidopsis thalia
43	42	39.3	156	AAW05861	Arabidopsis thalia
44	42	39.3	156	AAW53429	Arabidopsis thalia
45	42	39.3	276	AAW05860	Arabidopsis thalia

## ALIGNMENTS

RESULT 1  
AAV03708  
ID AAV03708 standard; Protein; 20 AA.  
XX  
AC AAV03708;  
XX  
DT 07-JUN-1999 (first entry)  
XX  
DE M. tuberculosis LHP polypeptide antigenic fragment.  
XX  
ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
KW immune response.

XX Mycobacterium tuberculosis.  
OS  
XX WO9904005-A1.  
PN  
XX 28-JAN-1999.  
PD  
XX 16-JUL-1998; 98WO-IB01091.  
XX  
XX 16-JUL-1997; 97US-0052631.  
PR  
XX (INSP ) INST PASTEUR.  
PA (STAT-) STAFENS SERUM INST.  
XX  
PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;  
XX WPI; 1999-132249/11.  
DR  
XX New nucleic acid containing regulator and LHP gene of Mycobacterium  
PT tuberculosis - useful in vaccines, for diagnosis, and for expression  
PT of heterologous proteins

XX Claim 21; Page 64; 88pp; English.

XX The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX

SQ Sequence 20 AA;

Query Match 83.68; Score 89.5; DB 20; Length 20;

Best Local Similarity 95.2%; Pred. No. 5e-08;

Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QEAGNFERISGLKVTQIDQV 21

|||||

Db 1 qeagnferisgdlk-tqidqv 20

RESULT 2

AAY03706

ID AAY03706 standard; Protein; 49 AA.

XX

AC AAY03706;

XX

DT 07-JUN-1999 (first entry)

XX

DE M. tuberculosis LHP polypeptide antigenic fragment.

XX

KW ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;

KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;

KW immune response.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9904005-A1.

XX

XX 28-JAN-1999.

PD

PF 16-JUL-1998; 98WO-IB01091.

XX

PR 16-JUL-1997; 97US-0052631.

XX

PA (INSP) INST PASTEUR.

XX

PA (STAT-) STATENS SERUM INST.

XX

PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX

XX WPI; 1999-132249/11.

DR

XX New nucleic acid containing regulator and LHP gene of Mycobacterium

PT tuberculosis - useful in vaccines, for diagnosis, and for expression

PT of heterologous proteins

XX

PS Claim 21; Page 64; 88pp; English.

XX

CC The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX

SQ Sequence 49 AA;

Query Match 83.68; Score 89.5; DB 20; Length 49;

Best Local Similarity 95.2%; Pred. No. 1.3e-07;

Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QEAGNFERISGLKVTQIDQV 21

|||||

Db 13 qeagnferisgdlk-tqidqv 32

RESULT 3

AAW32444

ID AAW32444 standard; Protein; 95 AA.

XX

AC AAW32444;

XX

DT 09-JAN-1998 (first entry)

XX

DE Mycobacterium tuberculosis antigen Tb38-1.

XX

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M. tuberculosis.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9709428-A2.

XX

PD 13-MAR-1997.

XX

PF 30-AUG-1996; 96WO-US14674.

XX

PR 12-JUL-1996; 96US-0680574.

PR 01-SEP-1995; 95US-0523436.

PR 22-SEP-1995; 95US-0533634.

PR 22-MAR-1996; 96US-0620874.

PR 05-JUN-1996; 96US-0659683.

XX

PA (CORI-) CORIXA CORP.

XX

PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

XX

DR WPI; 1997-192903/17.

DR N-PSDB; AAT91509.

XX

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are

PT useful in vaccines for prevention or treatment of tuberculosis, also

PT for diagnosis

XX

PS Example 3; Page 124; 168pp; English.

XX

CC A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or

CC its variant differing only in conservative substitutions and/or

CC modifications). The present sequence represents a M. tuberculosis

CC antigen, Tb38-1. The immunogenic protein, and fusion proteins

CC containing one or more of the proteins or one of the proteins plus

CC ESAT-6, are useful in vaccines, preferably when formulated with a

CC non-specific adjuvant, to induce an immune response against  
 CC M.tuberculosis (for treatment or prevention).  
 XX  
 SQ Sequence 95 AA;

Query Match 83.6%; Score 89.5; DB 18; Length 95;  
 Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QEAGNERISGDLKYTQIDQV 21  
 Db 8 qeagnerisgdlk-tqidqv 27

RESULT 4  
 AAW32376  
 ID AAW32376 standard; Protein; 95 AA.

AC AAW32376;  
 DT 13-JAN-1998 (first entry)  
 XX Mycobacterium tuberculosis antigen Tb38-1.

DE Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M.tuberculosis.  
 XX Mycobacterium tuberculosis.

OS  
 XX WO9709429-A2.  
 PN 13-MAR-1997.

PD 30-AUG-1996; 96WO-US14675.  
 PF 12-JUL-1996; 96US-0680573.  
 PR 01-SEP-1995; 95US-0523435.  
 PR 22-SEP-1995; 95US-0532136.  
 PR 22-MAR-1996; 96US-0620280.  
 PR 05-JUN-1996; 96US-0658800.

XX (CORI-) CORIXA CORP.  
 XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;  
 XX WPI; 1997-192904/17.  
 DR N-PSDB; AAT91445.

XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 - useful for diagnosis of M. tuberculosis infection

Example 3; Page 136; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M.tuberculosis  
 CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose  
 CC M.tuberculosis infection by forming complexes with specific  
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
 CC polypeptide can be used as diagnostic primers or probes and agents  
 CC that bind to the antigen, especially monoclonal antibodies or  
 CC equivalent polyclonal antibodies, are also used for diagnosis.

XX Sequence 95 AA;

Query Match 83.6%; Score 89.5; DB 18; Length 95;  
 Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QEAGNERISGDLKYTQIDQV 21  
 Db 8 qeagnerisgdlk-tqidqv 27

RESULT 5  
 AAW81747  
 ID AAW81747 standard; Protein; 95 AA.

XX AAW81747;  
 DT 27-JAN-1999 (first entry)  
 XX M. tuberculosis immunogenic polypeptide Tb38-1.

DE Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.  
 OS WO9816646-A2.  
 PN 23-APR-1998.

PD 07-OCT-1997; 97WO-US18293.  
 PF 13-MAR-1997; 97US-0818112.  
 PR 11-OCT-1996; 96US-0730510.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX WPI; 1998-261042/23.  
 DR N-PSDB; AAV64491.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis

PS Example 3b; Page 117; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This sequence  
 CC can be formulated into vaccines and/or pharmaceutical compositions for  
 CC immunising against M. tuberculosis infection or may be used for the  
 CC diagnosis of tuberculosis.

SQ Sequence 95 AA;

Query Match 83.6%; Score 89.5; DB 19; Length 95;  
 Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QEAGNERISGDLKYTQIDQV 21  
 Db 8 qeagnerisgdlk-tqidqv 27

RESULT 6  
 AAW64321  
 ID AAW64321 standard; Peptide; 95 AA.

XX AAW64321;  
 DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb38-1 peptide.  
 XX Tuberculosis; infection; diagnosis; antigen; Tb38-1.

XX OS Mycobacterium tuberculosis strain H37Rv.  
 XX PN WO9816645-A2.  
 XX PD 23-APR-1998.  
 XX PF 07-OCT-1997; 97WO-US18214.  
 XX PR 13-MAR-1997; 97US-0818111.  
 XX PR 11-OCT-1996; 96US-0729622.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX DR WPI; 1998-251292/22.  
 XX DR N-PSDB; AAV44384.  
 XX PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 XX PT to develop products for the detection of M. tuberculosis infection  
 XX PT and diagnosis of tuberculosis  
 XX PS Example 3; Page 123; 250pp; English.  
 XX CC This is an antigenic portion of Mycobacterium tuberculosis antigen  
 XX CC Tb38-1. A DNA sequence (see AAV44384) coding for antigen Tb38-1 was  
 XX CC isolated from a M. tuberculosis strain H37Rv expression library  
 XX CC using sera from patients having pulmonary or pleural tuberculosis.  
 XX CC The invention relates to compositions and methods for diagnosing  
 XX CC tuberculosis. It provides polypeptides (see AAV44291-W64379)  
 XX CC comprising an antigenic portion of a soluble M. tuberculosis  
 XX CC antigen, or an immunogenic portion of an M. tuberculosis antigen,  
 XX CC as well as DNA sequences encoding such polypeptides, recombinant  
 XX CC expression vectors and transformed or transfected host cells. Also  
 XX CC claimed are methods and diagnostic kits for detecting M.  
 XX CC tuberculosis infection in a patient using these polypeptides,  
 XX CC antibodies or oligonucleotide probes and primers, for the diagnosis  
 XX CC of tuberculosis.  
 XX SQ Sequence 95 AA;

Query Match 83.6%; Score 89.5; DB 19; Length 95;  
 Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 QEAGNFERISGDLKYTOIDQV 21  
 Db 8 qeagnferisgdlk-tqidqv 27  
 |||||

RESULT 7  
 AAY32097  
 ID AAY32097 standard; Protein; 95 AA.  
 XX AC  
 XX AAY32097;  
 XX DT 17-JAN-2000 (first entry)  
 XX DE Mycobacterium tuberculosis antigen Tb38-1.  
 XX KW Tuberculosis; antigen; fusion protein; Tb38-1; diagnosis; therapy;  
 XX KW vaccine; immunogen.  
 XX OS Mycobacterium tuberculosis.  
 XX PN WO9951748-A2.  
 XX PD 14-OCT-1999.  
 XX PF 07-APR-1999; 99WO-US07717.

XX 07-APR-1998; 98US-0056556.  
 XX 30-DEC-1998; 98US-0223040.  
 XX PA (CORI-) CORIXA CORP.  
 XX XX Skeiky YAW, Alderson M, Campos-Neto A;  
 XX PI WPI; 1999-601610/51.  
 XX DR  
 XX XX New fusion proteins useful for diagnosis, prevention and treatment of  
 XX PT tuberculosis -  
 XX XX Claim 1; Fig 4D; 83pp; English.  
 XX XX This sequence represents the Mycobacterium tuberculosis antigen  
 XX CC Tb38-1. The invention provides fusion proteins (see AAY32059-71)  
 XX CC containing at least 2 M. tuberculosis antigens such as Tb38-1, e.g.  
 XX CC Tbf-2 (see AAY32063) and a Tbf9-Tb38-1 fusion. The new fusion  
 XX CC proteins are useful as vaccines for preventing tuberculosis  
 XX CC (claimed), for diagnosis (via in vitro assays or intradermal skin  
 XX CC tests for detection of anti-M. tuberculosis antibodies), monitoring  
 XX CC of disease progression, and treatment of tuberculosis. They are  
 XX CC more effective immunogens than mixtures of the individual protein  
 XX CC components.  
 XX SQ Sequence 95 AA;

Query Match 83.6%; Score 89.5; DB 20; Length 95;  
 Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 QEAGNFERISGDLKYTOIDQV 21  
 Db 8 qeagnferisgdlk-tqidqv 27  
 |||||

RESULT 8  
 AAY39118  
 ID AAY39118 standard; Protein; 95 AA.  
 XX AC  
 XX AAY39118;  
 XX DT 05-NOV-1999 (first entry)  
 XX DE M. tuberculosis antigen Tb38-1 amino acid sequence.  
 XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 XX KW immune response; skin test.  
 XX OS Mycobacterium tuberculosis.  
 XX PN WO9942076-A2.  
 XX PD 26-AUG-1999.  
 XX PF 17-FEB-1999; 99WO-US03268.  
 XX PR 05-MAY-1998; 98US-0072967.  
 XX PR 18-FEB-1998; 98US-0025197.  
 XX XX (CORI-) CORIXA CORP.  
 XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX DR WPI; 1999-527409/44.  
 XX XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
 XX PT skin tests and protective or therapeutic vaccines or compositions  
 XX XX

PS Example 3; Page 113; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AA219249 to AA219460 and AA39083 to  
CC AA39225 are used in the exemplification of the present invention.

XX Sequence 95 AA;

Query Match 83.6%; Score 89.5; DB 20; Length 95;  
Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QEAGNFERISGDLKYTQIDQV 21  
Db 8 qeagnferisgdlk-tqidqv 27

RESULT 9  
AA38981  
ID AAY38981 standard; Protein; 95 AA.

XX AAY38981;

XX 05-NOV-1999 (first entry)

XX M. tuberculosis recombinant antigen protein Tb38-1.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;  
KW vaccine; immunity.

OS Mycobacterium tuberculosis.

XX WO9942118-A2.

PN 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03265.

XX 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527416/44.

DR N-PSDB; AA219082.

XX New polypeptide comprising antigenic portions of M. tuberculosis

PS Example 3; Page 159; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding  
CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
CC polypeptides are useful for detecting M. tuberculosis infection in a  
CC biological sample by detecting antibodies which bind with the  
CC polypeptides, and are useful as vaccines for immunizing against  
CC M. tuberculosis infection. The new detection methods are needed as  
CC current vaccination strategies do not provide 100% immunity.

XX Sequence 95 AA;

Query Match 83.6%; Score 89.5; DB 20; Length 95;  
Best Local Similarity 95.2%; Pred. No. 2.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QEAGNFERISGDLKYTQIDQV 21  
Db 8 qeagnferisgdlk-tqidqv 27

RESULT 10  
AAW81706  
ID AAW81706 standard; Protein; 100 AA.

XX AAW81706;

XX 27-JAN-1999 (first entry)

XX M. tuberculosis immunogenic polypeptide Tb38-IN.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

OS WO9816646-A2.

PN 23-APR-1998.

XX 07-OCT-1997; 97WO-US18293.

XX 13-MAR-1997; 97US-0818112.

PR 11-OCT-1996; 96US-0730510.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis

XX Example 3B; Page 138-139; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble  
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
CC for inducing protective immunity against tuberculosis (TB). This  
CC sequence can be formulated into vaccines and/or pharmaceutical  
CC compositions for immunising against M. tuberculosis infection or may  
CC be used for the diagnosis of tuberculosis.

XX Sequence 100 AA;

Query Match 83.6%; Score 89.5; DB 19; Length 100;  
Best Local Similarity 95.2%; Pred. No. 2.9e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 QEAGNFERISGDLKYTQIDQV 21  
Db 13 qeagnferisgdlk-tqidqv 32

RESULT 11  
AAW64339  
ID AAW64339 standard; Protein; 100 AA.

XX AAW64339;

XX 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb38-IN.  
 XX Tuberculosis; infection; diagnosis; antigen; Tb37-FL.  
 XX OS Mycobacterium tuberculosis strain H37Rv.  
 XX PN WO9816645-A2.  
 XX PD 23-APR-1998.  
 XX PF 07-OCT-1997; 97WO-US18214.  
 XX PR 13-MAR-1997; 97US-0818111.  
 XX PR 11-OCT-1996; 96US-0729622.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX DR WPI; 1998-251292/22.  
 XX PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 XX PT to develop products for the detection of M. tuberculosis infection  
 XX PT and diagnosis of tuberculosis  
 XX PS Example 3; Page 145; 250pp; English.  
 XX CC This polypeptide comprises a partial sequence of Mycobacterium  
 XX CC tuberculosis antigen Tb38-IN. It is encoded by genomic DNA isolated  
 XX CC from a M. tuberculosis strain H37Rv genomic library using a probe  
 XX CC derived from clone Tb38-1 (see AAV44384). The invention relates to  
 XX CC compositions and methods for diagnosing tuberculosis. It provides  
 XX CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of  
 XX CC a soluble M. tuberculosis antigen, or an immunogenic portion of an  
 XX CC M. tuberculosis antigen, as well as DNA sequences encoding such  
 XX CC polypeptides, recombinant expression vectors and transformed or  
 XX CC transfectant host cells. Also claimed are methods and diagnostic  
 XX CC kits for detecting M. tuberculosis infection in a patient using  
 XX CC these polypeptides, antibodies or oligonucleotide probes and  
 XX CC primers, for the diagnosis of tuberculosis.  
 XX SQ Sequence 100 AA;  
 Query Match 83.6%; Score 89.5; DB 19; Length 100;  
 Best Local Similarity 95.2%; Pred. No. 2.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 Qy 1 QEAGNFERISGDLKYTDQV 21  
 Db 13 qeagnferisgdlk-tqidv 32  
 RESULT 12  
 AAY39136  
 ID AAY39136 standard; Protein; 100 AA.  
 XX AC AAY39136;  
 XX DT 05-NOV-1999 (first entry)  
 XX DE M. tuberculosis antigen Tb38-IN amino acid sequence.  
 XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 XX KW immune response; skin test.  
 XX OS Mycobacterium tuberculosis.  
 XX PN WO9942076-A2.  
 XX PD 26-AUG-1999.  
 Query Match 83.6%; Score 89.5; DB 20; Length 100;  
 Best Local Similarity 95.2%; Pred. No. 2.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 Qy 1 QEAGNFERISGDLKYTDQV 21  
 Db 13 qeagnferisgdlk-tqidv 32  
 RESULT 13  
 AAY38993  
 ID AAY38993 standard; Protein; 100 AA.  
 XX AC AAY38993;  
 XX DT 05-NOV-1999 (first entry)  
 XX DE M. tuberculosis recombinant antigen protein Tb38-IN.  
 XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
 XX KW vaccine; immunity.  
 XX OS Mycobacterium tuberculosis.  
 XX PN WO9942118-A2.  
 XX PD 26-AUG-1999.  
 XX PF 17-FEB-1999; 99WO-US03265.  
 XX PR 05-MAY-1998; 98US-0072596.  
 XX PR 18-FEB-1998; 98US-0024753.  
 XX PA (CORI-) CORIXA CORP.  
 XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX DR WPI; 1999-527416/44.  
 XX PD



PT New polypeptide comprising antigenic portions of M. tuberculosis  
 PS Example 3; Page 179; 323pp; English.  
 XX This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.  
 XX  
 SQ Sequence 100 AA;  
 Query Match 83.6%; Score 89.5; DB 20; Length 100;  
 Best Local Similarity 95.2%; Pred. No. 2.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 QY 1 QEAGNFERISGDLKYTQIDQV 21  
 |||||  
 Db 13 qeagnferisgdlk-tqidqv 32  
 RESULT 14  
 AAY03705  
 ID AAY03705 standard; Protein; 100 AA.  
 AC AAY03705;  
 XX  
 XX 07-JUN-1999 (first entry)  
 DT M. tuberculosis LHP polypeptide.  
 DE  
 XX  
 XX ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
 KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
 KW immune response.  
 XX  
 XX Mycobacterium tuberculosis.  
 OS  
 XX  
 XX WO9904005-A1.  
 PN  
 XX  
 XX 28-JAN-1999.  
 PD  
 XX  
 XX 16-JUL-1998; 98WO-TB01091.  
 PF  
 XX  
 XX 16-JUL-1997; 97US-0052631.  
 PR  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (STAT-) STATENS SERUM INST.  
 XX  
 XX Andersen P, Berthet F, Gicquel B, Rasmussen PB;  
 PI  
 XX WPI; 1999-132249/11.  
 DR  
 DR N-PSDB; AAX291168, AAX291171.  
 XX  
 XX New nucleic acid containing regulator and LHP gene of Mycobacterium  
 PT tuberculosis - useful in vaccines, for diagnosis, and for expression  
 PT of heterologous proteins  
 XX  
 XX Claim 17; Page 64; 88pp; English.  
 PS  
 XX  
 XX The present invention is directed to a polynucleotide carrying the  
 CC regulatory expression signals of the ESAT-6 protein as well as an open  
 CC reading frame coding for an antigenic protein LHP from Mycobacterium  
 CC tuberculosis. Host cells comprising the polynucleotide are used for the  
 CC recombinant expression of the protein. The recombinant polypeptide can be  
 CC used as immunogens and vaccines, to protect against bacteria of the  
 CC M. tuberculosis complex in humans or animals (the vaccines may include  
 CC other immunogenic proteins of the bacteria or their fragments,  
 CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by  
 CC detection of specific antibodies. The regulatory region present in the  
 CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two  
 CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to  
 CC provide a synergistic increase in ability to induce a protective immune  
 CC response. The present sequence represents the LHP polypeptide.  
 XX  
 SQ Sequence 100 AA;  
 Query Match 83.6%; Score 89.5; DB 20; Length 100;  
 Best Local Similarity 95.2%; Pred. No. 2.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 QY 1 QEAGNFERISGDLKYTQIDQV 21  
 |||||  
 Db 13 qeagnferisgdlk-tqidqv 32  
 RESULT 15  
 AAB35218  
 ID AAB35218 standard; Protein; 100 AA.  
 XX  
 AC AAB35218;  
 XX  
 XX 24-APR-2001 (first entry)  
 DT M tuberculosis Rv3874 protein.  
 DE  
 XX  
 XX Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv10366;  
 KW Rv1037C; Rv2348C; Rv2853C; Rv2854C; Rv3020C; Rv3444C;  
 KW Rv3445C; Rv3890C; Rv3891C; Rv3904C; Rv3905C.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 XX WO200104151-A2.  
 PN  
 XX  
 XX 18-JAN-2001.  
 PD  
 XX  
 XX 13-JUL-2000; 2000WO-DK00398.  
 PF  
 XX  
 XX 13-JUL-1999; 99DK-0001020.  
 PR  
 XX 15-JUL-1999; 99US-0144011.  
 PR  
 XX  
 XX (STAT-) STATENS SERUM INST.  
 PA  
 XX  
 XX Andersen P, Skjot R;  
 PI  
 XX WPI; 2001-091923/10.  
 DR  
 XX  
 XX New polypeptide encoded by a member of the esat-6-gene family for  
 PT immunizing against and diagnosis of tuberculosis -  
 PT  
 XX  
 XX Example 2; Page 65; 80pp; English.  
 PS  
 XX  
 XX The present invention provides the protein and coding sequences for  
 CC members of the esat-6 gene family from Mycobacterium tuberculosis. These  
 CC proteins include Rv0287, Rv1036C, Rv1037C, Rv2348C, Rv2653C,  
 CC Rv2654C, Rv3020C, Rv3444C, Rv3445C, Rv3890C, Rv3891C, Rv3904C and  
 CC Rv3905C. These can be used to produce vaccines against, and in the  
 CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of  
 CC the proteins of the invention.  
 XX  
 SQ Sequence 100 AA;  
 Query Match 83.6%; Score 89.5; DB 22; Length 100;  
 Best Local Similarity 95.2%; Pred. No. 2.9e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 QY 1 QEAGNFERISGDLKYTQIDQV 21  
 |||||  
 Db 13 qeagnferisgdlk-tqidqv 32

Search completed: July 5, 2001, 11:45:43  
Job time: 272 sec

3

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:46:59 ; Search time 61.79 Seconds  
(without alignments)  
6.846 Million cell updates/sec

Title: US-09-462-480-8  
Perfect score: 107  
Sequence: 1 QEAGNFERISGLKYTQIDQV 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2.6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2.6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2.6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2.6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2.6/ptodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46.5	43.5	360	1 US-08-317-223-2	Sequence 2, Appli
2	46.5	43.5	360	1 US-08-445-135-3	Sequence 3, Appli
3	46.5	43.5	360	3 US-09-059-849A-2	Sequence 2, Appli
4	46.5	43.5	360	4 US-09-213-632-2	Sequence 2, Appli
5	46.5	43.5	360	5 PCT-US95-12675-2	Sequence 2, Appli
6	42	39.3	831	2 US-08-677-734A-11	Sequence 11, Appl
7	41	38.3	141	1 US-08-470-179-17	Sequence 17, Appl
8	41	38.3	141	1 US-08-470-179-18	Sequence 18, Appl
9	41	38.3	141	1 US-08-470-179-19	Sequence 19, Appl
10	41	38.3	828	1 US-08-261-304-2	Sequence 2, Appli
11	41	38.3	1290	1 US-08-138-641-2	Sequence 2, Appli
12	41	38.3	1290	1 US-08-138-133-2	Sequence 2, Appli
13	40	37.4	141	1 US-08-470-179-20	Sequence 20, Appl
14	39.5	36.9	540	1 US-08-367-227-2	Sequence 2, Appli
15	39	36.4	140	1 US-08-470-179-21	Sequence 21, Appl
16	39	36.4	377	1 US-08-525-697-2	Sequence 2, Appli
17	39	36.4	425	1 US-08-700-749A-5	Sequence 5, Appli
18	39	36.4	425	3 US-09-020-467-5	Sequence 5, Appli
19	39	36.4	425	3 US-09-020-467-5	Sequence 5, Appli
20	39	36.4	425	3 US-09-020-685-5	Sequence 5, Appli
21	39	36.4	800	6 5183745-3	Patent No. 5183745
22	39	36.4	1229	1 US-08-100-709-4	Sequence 4, Appli
23	39	36.4	1229	1 US-08-176-865-4	Sequence 4, Appli
24	39	36.4	1229	1 US-08-474-038-4	Sequence 4, Appli
25	39	36.4	1229	2 US-08-779-046-4	Sequence 4, Appli
26	39	36.4	1229	2 US-08-881-340-4	Sequence 4, Appli
27	38	35.5	677	4 US-08-836-567-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-317-223-2  
Sequence 2, Application US/08317223  
Patent No. 5585267  
GENERAL INFORMATION:  
APPLICANT: Jones, Jonathan C.R.  
APPLICANT: Quaranta, Vito  
APPLICANT: Tamura, Richard  
TITLE OF INVENTION: CELLULAR ATTACHMENT TO TRANS-EPITHELIAL  
TITLE OF INVENTION: APPLIANCES  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/317,223  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,727  
FILING DATE: 05-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/151,134  
FILING DATE: 12-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/152,460  
FILING DATE: 12-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: DESMOS.002CP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

28 38 35.5 1197 4 US-08-836-567-12 Sequence 12, Appli  
29 38 35.5 1230 2 US-08-968-542C-35 Sequence 35, Appli  
30 38 35.5 1551 4 US-09-425-665-2 Sequence 2, Appli  
31 37 34.6 141 1 US-08-470-179-25 Sequence 25, Appli  
32 37 34.6 221 1 US-08-393-058-1 Sequence 1, Appli  
33 37 34.6 707 4 US-08-704-711A-19 Sequence 19, Appli  
34 37 34.6 708 4 US-08-448-489-16 Sequence 16, Appli  
35 36.5 34.1 358 1 US-08-700-186-2 Sequence 2, Appli  
36 36.5 34.1 358 2 US-08-914-981-2 Sequence 2, Appli  
37 36.5 34.1 358 3 US-09-116-115-2 Sequence 2, Appli  
38 36.5 34.1 436 3 US-08-669-378-8 Sequence 8, Appli  
39 36 33.6 141 1 US-08-470-179-24 Sequence 24, Appli  
40 36 33.6 241 1 US-08-212-190A-2 Sequence 2, Appli  
41 36 33.6 241 2 US-08-900-321-2 Sequence 2, Appli  
42 36 33.6 241 5 PCT-US95-03610-2 Sequence 2, Appli  
43 36 33.6 397 4 US-09-193-191-2 Sequence 2, Appli  
44 36 33.6 402 2 US-08-394-189B-14 Sequence 14, Appli  
45 36 33.6 402 3 US-08-368-704C-38 Sequence 38, Appli

;  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; IMMEDIATE SOURCE:  
; CLONE: laminin A  
US-08-317-223-2

Query Match 43.5%; Score 46.5; DB 1; Length 360;  
Best Local Similarity 55.0%; Pred. No. 3.4;  
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 QEAGNFERISGLKYTQIDQ 20  
:||||| :||||| :|:|  
Db 17 KDAGNF-LINGDLTLNQINQ 35

RESULT 2  
US-08-445-135-3  
; Sequence 3, Application US/08445135  
; Patent No. 5658789  
; GENERAL INFORMATION:  
; APPLICANT: Quaranta, Vito  
; APPLICANT: Hornia, Marketa  
; TITLE OF INVENTION: Promotion of Epithelial Cell Adhesion  
; and Hemidesmosome Assembly by a Laminin-like Molecule  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson and Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445.135  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/151,134  
; FILING DATE: 12-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: DESMOS.002CP2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 360 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; IMMEDIATE SOURCE:  
; CLONE: laminin A  
US-08-445-135-3

Query Match 43.5%; Score 46.5; DB 1; Length 360;  
Best Local Similarity 55.0%; Pred. No. 3.4;  
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 QEAGNFERISGLKYTQIDQ 20

Db 17 KDAGNF-LINGDLTLNQINQ 35  
:||||| :||||| :|:|

RESULT 3  
US-09-059-849A-2  
; Sequence 2, Application US/09059849A  
; Patent No. 6034068  
; GENERAL INFORMATION:  
; APPLICANT: Craig Halberstadt  
; TITLE OF INVENTION: CELLULAR ATTACHMENT TO LAMININ 5-COATED  
; TRANS-EPITHELIAL APPLIANCES  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor  
; CITY: Newport Beach  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/059,849A  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelsen, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: DESMOS.015A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 360 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; IMMEDIATE SOURCE:  
; CLONE: laminin A  
US-09-059-849A-2

Query Match 43.5%; Score 46.5; DB 3; Length 360;  
Best Local Similarity 55.0%; Pred. No. 3.4;  
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 QEAGNFERISGLKYTQIDQ 20  
:||||| :||||| :|:|  
Db 17 KDAGNF-LINGDLTLNQINQ 35

RESULT 4  
US-09-213-632-2  
; Sequence 2, Application US/09213632  
; Patent No. 6110711  
; GENERAL INFORMATION:  
; APPLICANT: Jones, Jonathan C.R.  
; APPLICANT: Quaranta, Vito  
; APPLICANT: Tamura, Richard  
; TITLE OF INVENTION: CELLULAR ATTACHMENT TO TRANS-EPITHELIAL  
; APPLIANCES

NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/213,632  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/317,223  
FILING DATE:  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/151,134  
FILING DATE: 12-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/152,460  
FILING DATE: 12-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: DESMOS.002CPI  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
IMMEDIATE SOURCE:  
CLONE: laminin A  
US-09-213-632-2

Query Match 43.5%; Score 46.5; DB 4; Length 360;  
Best Local Similarity 55.0%; Pred. No. 3.4;  
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 QEAGNERISGDLKYQTIDQ 20  
Db :|||||:|||||:|

Db 17 KDAGNF-LINGDLTLNQIQ 35

RESULT 5  
PCT-US95-12675-2  
Sequence 2, Application PC/TUS9512675  
GENERAL INFORMATION:  
APPLICANT: DESMOS, Inc.  
TITLE OF INVENTION: CELLULAR ATTACHMENT TO TRANS-EPITHELIAL  
TITLE OF INVENTION: APPLIANCES  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/12675  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/317,223  
FILING DATE: 03-OCT-1994  
APPLICATION NUMBER: US 08/042,727  
FILING DATE: 05-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/151,134  
FILING DATE: 12-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/152,460  
FILING DATE: 12-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: DESMOS.002CPI  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
IMMEDIATE SOURCE:  
CLONE: laminin A  
PCT-US95-12675-2

Query Match 43.5%; Score 46.5; DB 5; Length 360;  
Best Local Similarity 55.0%; Pred. No. 3.4;  
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 QEAGNERISGDLKYQTIDQ 20  
Db :|||||:|||||:|

Db 17 KDAGNF-LINGDLTLNQIQ 35

RESULT 6  
US-08-677-734A-11  
Sequence 11, Application US/08677734A  
Patent No. 5871919

GENERAL INFORMATION:  
APPLICANT: Brant, Steven R.  
APPLICANT: Yun, Chris C.H.  
APPLICANT: Donowitz, Mark  
APPLICANT: Tse, Chung-Ming  
TITLE OF INVENTION: Cloning, Tissue Distribution, and  
TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/677,734A  
FILING DATE: 10-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fordis, Jean B.  
REGISTRATION NUMBER: 32,984  
REFERENCE/DOCKET NUMBER: 05387.0043-00000  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 831 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-677-734A-11

Query Match 39.3%; Score 42; DB 2; Length 831;  
Best Local Similarity 46.7%; Pred. No. 49;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 QEAGNFERISGLKV 15  
:|||:||||:|:  
Db 719 EATNYEISGIEF 733

RESULT 7  
US-08-470-179-17  
; Sequence 17, Application US/08470179  
; Patent No. 5645994  
; GENERAL INFORMATION:  
; APPLICANT: Huang Ph.D, Wai Mun  
; TITLE OF INVENTION: Method and Compositions for  
; IDENTIFICATION OF SPECIES IN A SAMPLE  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask, Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,179  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert Ph.D, Susan E.  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2601  
; TELEPHONE: 801-532-1922  
; TELEFAX: 801-531-9168  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 141 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO

; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
US-08-470-179-17

Query Match 38.3%; Score 41; DB 1; Length 141;  
Best Local Similarity 47.4%; Pred. No. 9.8;  
Matches 9; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 4 GNERISGD----LKVTQI 18  
||| | | :|||:  
Db 69 GNFGSIDGDSAAARMYTEI 87

RESULT 8  
US-08-470-179-18  
; Sequence 18, Application US/08470179  
; Patent No. 5645994  
; GENERAL INFORMATION:  
; APPLICANT: Huang Ph.D, Wai Mun  
; TITLE OF INVENTION: Method and Compositions for  
; IDENTIFICATION OF SPECIES IN A SAMPLE  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask, Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,179  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert Ph.D, Susan E.  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2601  
; TELEPHONE: 801-532-1922  
; TELEFAX: 801-531-9168  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 141 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Shigella dysenteriae  
US-08-470-179-18

Query Match 38.3%; Score 41; DB 1; Length 141;  
Best Local Similarity 47.4%; Pred. No. 9.8;  
Matches 9; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 4 GNERISGD----LKVTQI 18  
||| | | :|||:  
Db 69 GNFGSIDGDSAAARMYTEI 87

RESULT 9

US-08-470-179-19  
; Sequence 19, Application US/08470179  
; Patent No. 5645994  
; GENERAL INFORMATION:  
; APPLICANT: Huang Ph.D, Wai Mun  
; TITLE OF INVENTION: Method and Compositions for  
; IDENTIFICATION OF SPECIES IN A SAMPLE  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask, Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,179  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert Ph.D, Susan E.  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2601  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 801-532-1922  
; TELEFAX: 801-531-9168  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 141 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE:  
; ORGANISM: Salmonella typhimurium  
US-08-470-179-19

Query Match 38.3%; Score 41; DB 1; Length 141;  
Best Local Similarity 47.4%; Pred. No. 9.8;  
Matches 9; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 4 GNFERISGD---LKYTOI 18  
||| ||| :|||  
Db 69 GNFGSIDGSAAMRYTEI 87

RESULT 10  
US-08-261-304-2  
; Sequence 2, Application US/08261304  
; Patent No. 5708147  
; GENERAL INFORMATION:  
; APPLICANT: Cybulsky, Myron I.  
; APPLICANT: Gimbrone, Michael A.  
; APPLICANT: Collins, Tucker  
; TITLE OF INVENTION: Mononuclear Leukocyte Directed  
; TITLE OF INVENTION: Endothelial Adhesion Molecule Associated with  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1225 Connecticut Avenue, N.W.  
; CITY: Suite 300  
; STATE: Washington  
; DISTRICT OF Columbia

COUNTRY: United States of America  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Ascii  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/261,304  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/649,565  
FILING DATE: 01-FEB-1991  
APPLICATION NUMBER: U.S. 07/487,038  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E. R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 0627.2100004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 833-7533  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 828 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-261-304-2

Query Match 38.3%; Score 41; DB 1; Length 828;  
Best Local Similarity 44.4%; Pred. No. 72;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 QEAGNFERISGDLKYTOI 18  
|: ||: ||: ||  
Db 262 QDNGNLQRLSGNATLTI 279

RESULT 11  
US-08-138-641-2  
; Sequence 2, Application US/08138641  
; Patent No. 5474921  
; GENERAL INFORMATION:  
; APPLICANT: Koblan, Kenneth S.  
; APPLICANT: Pomplano, David L.  
; TITLE OF INVENTION: ASSAY TO DETERMINE INHIBITORS OF  
; PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C-GAMMA  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Muthard  
; STREET: P.O. Box 2000, 126 E. Lincoln Avenue  
; CITY: Rahway  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07065  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/138,641  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Muthard, David A.  
; REGISTRATION NUMBER: 35,297  
; REFERENCE/DOCKET NUMBER: 18937  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 594-3903  
; TELEFAX: (908) 594-4720

```

; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1290 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
; US-08-138-641-2

```

Query Match 38.3%; Score 41; DB 1; Length 1290;  
Best Local Similarity 47.1%; Pred. NO. 1.2e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels

RESULT 12  
US-08-138-133-2  
; Sequence 2, Application US/08138133  
: Patent No 5519163

Query Match 38.3%; Score 41; DB 1; Length 1290;  
Best Local Similarity 47.1%; Pred. No. 1.2e+02;  
Matches 8; Conservative 4; Mismatches 5; Indels

RESULT 13  
US-08-470-179-20  
; Sequence 20, Application US/08470179  
; Patent No. 5645994  
; GENERAL INFORMATION:  
; APPLICANT: Huang Ph.D, Wai Mun  
; TITLE OF INVENTION: Method and Compositions for  
; IDENTIFICATION OF SPECIES IN A SAMPLE  
; NUMBER OF SEQUENCES: 207  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Trask, Britt and Rossa  
; STREET: P.O. Box 2550  
; CITY: Salt Lake City  
; STATE: Utah  
; COUNTRY: USA  
; ZIP: 84110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,179  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sweigert Ph.D, Susan E.  
; REGISTRATION NUMBER: 36,289  
; REFERENCE/DOCKET NUMBER: 2601  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 801-532-1922  
; TELEFAX: 801-531-9168  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 141 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Klebsiella pneumoniae  
; US-08-470-179-20

RESULT 14  
US-08-367-227-2  
; Sequence 2, Application US/08367227  
; Patent No. 5587304  
; GENERAL INFORMATION:  
; APPLICANT: BARRE, PIERRE  
; APPLICANT: DEQUIN, SYLVIE  
; APPLICANT: ANSANAY, VIRGINIE  
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THE GENE FOR  
; TITLE OF INVENTION: THE MALOLACTIC ENZYME OF LACTOCOCCUS LACTIS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA



ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/367,227  
FILING DATE: 17-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR PCT/FR94/00589  
FILING DATE: 18-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 93/06003  
FILING DATE: 18-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 384-42-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 540 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-367-227-2

Query Match 36.9%; Score 39.5; DB 1; Length 540;  
Best Local Similarity 55.6%; Pred. No. 80;  
Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 4 GNF-ERISGDLKYTDQ 20  
|| ||| ||| : |||  
Db 200 GNHQRIGDQYYSFVDQ 217

RESULT 15  
US-08-470-179-21  
Sequence 21, Application US/08470179  
Patent No. 5645994  
GENERAL INFORMATION:  
APPLICANT: Huang Ph.D, Wai Mun  
TITLE OF INVENTION: Method and Compositions for  
IDENTIFICATION OF SPECIES  
NUMBER OF SEQUENCES: 207  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Trask, Britt and Rossa  
STREET: P.O. Box 2550  
CITY: Salt Lake City  
STATE: Utah  
COUNTRY: USA  
ZIP: 84110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,179  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sweigert Ph.D, Susan E.  
REGISTRATION NUMBER: 36,289  
REFERENCE/DOCKET NUMBER: 2601  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 801-532-1922

TELEFAX: 801-531-9168  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 140 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas aeruginosa  
US-08-470-179-21

Query Match 36.4%; Score 39; DB 1; Length 140;  
Best Local Similarity 36.8%; Pred. No. 21;  
Matches 7; Conservative 5; Mismatches 3; Indels 4; Gaps

QY 4 GNFERISGD---LKYTQI 18  
||| : ||| : ||| : |||  
Db 68 GNFGSVGDGDNAAAMRYTEV 86

Search completed: July 5, 2001, 11:46:59  
Job time: 313 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:48:32 ; Search time 79.63 seconds  
(without alignments)  
20.089 Million cell updates/sec

Title: US-09-462-480-8  
Perfect score: 107

Sequence: 1 QEAGNFERISGDLKYTQIDQV 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	89.5	83.6	100	2	H70802	hypothetical prote
2	51	47.7	296	2	T27375	hypothetical prote
3	46.5	43.5	3712	2	S18253	laminin alpha-1 ch
4	46	43.0	297	2	F69595	multidrug resistan
5	46	43.0	547	2	S44841	K06H7.1 protein -
6	44	41.1	411	2	A48741	phosphopentomutase
7	43	40.2	317	2	F85650	hypothetical prote
8	43	40.2	643	2	G72320	penicillin-binding
9	43	40.2	909	2	C84965	oxoglutarate dehyd
10	43	40.2	948	2	JC2190	preprotein translo
11	43	40.2	1023	1	LEECA	hemolysin A - Esch
12	43	40.2	1279	2	A47363	RNA helicase A - h
13	43	40.2	1287	2	I46032	nuclear DNA helica
14	42.5	39.7	314	1	JC1276	dihydroorotate oxi
15	42.5	39.7	1014	1	S75023	sensory transducti
16	42	39.3	462	2	S55114	hypothetical prote
17	42	39.3	480	2	G75313	molybdate metaboli
18	42	39.3	540	2	T51390	TCP-1 chaperonin-1
19	42	39.3	702	2	A69140	ATP-dependent heli
20	42	39.3	831	2	B40204	Na+/H+-exchanging
21	42	39.3	859	2	T26134	hypothetical prote
22	42	39.3	1616	2	T47801	hypothetical prote
23	41.5	38.8	307	2	T47244	C-rat protein - ze
24	41.5	38.8	1439	2	T02087	gag/pol polyprotei
25	41	38.3	95	2	S54254	DNA topoisomerase
26	41	38.3	189	2	A38577	lumazine protein L
27	41	38.3	192	2	A64098	molybdopterin-guan
28	41	38.3	305	2	C83967	aspartate carbamoy
29	41	38.3	341	2	T26552	hypothetical prote

30 41 38.3 393 2 JC4318 inulin fructotrans  
31 41 38.3 451 2 G72217 replicative DNA he  
32 41 38.3 463 2 D72356 conserved hypot het  
33 41 38.3 875 1 ITECAP DNA topoisomerase  
34 41 38.3 875 2 D85862 DNA gyrase, subuni  
35 41 38.3 878 2 S47332 DNA topoisomerase  
36 41 38.3 1217 2 T21403 hypothetical prote  
37 41 38.3 1290 2 A31317 1-phosphatidylinos  
38 41 38.3 1511 2 S60932 probable membrane  
39 40.5 37.9 1028 2 A96719 hypothetical prote  
40 40 37.4 93 2 B86677 hypothetical prote  
41 40 37.4 95 2 S49447 DNA topoisomerase  
42 40 37.4 125 2 G75351 hypothetical prote  
43 40 37.4 148 2 T22886 hypothetical prote  
44 40 37.4 213 2 E83386 hypothetical prote  
45 40 37.4 213 2 T49938 hypothetical prote

ALIGNMENTS

RESULT 1

H70802

hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C:Accession: H70802

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:9825987

A:Accession: H70802

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-100 <COL>

A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL1966.1; PID:g296

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3874

Query Match 83.6%; Score 89.5; DB 2; Length 100;  
Best Local Similarity 95.2%; Pred. No. 1.6e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps

QY 1 QEAGNFERISGDLKYTQIDQV 21

|||||

Db 13 QEAGNFERISGDLK-TQIDQV 32

RESULT 2

T27375

hypothetical protein Y75B12B.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27375

R:White, S.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20360

A:Accession: T27375

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-296 <WIL>

A:Cross-references: EMBL:AL032663; PIDN:CAA21764.1; GSPDB:GN00023; CESP:Y75B12B.1

A:Experimental source: clone Y75B12B

C:Genetics:

A:Gene: CESP:Y75B12B.1

A:Map position: 5

A:Introns: 99/3; 173/2; 199/3; 248/3

[illegible]

C;Accession: A86741  
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich, G.; et al. 2001  
Genome Res. in press, 2001  
A;Title: The complete genome sequence of the lactic acid bacterium.  
A;Reference number: A86625  
A;Accession: A86741  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-411 <STO>  
A;Cross-references: GB:AE005176; NID:g12723864; PIDN:AAK05027.1; GSPDB:GN00146  
A;Experimental source: strain IL1403  
C;Genetics:  
A;Gene: deoB  
C;Superfamily: phosphopentomutase  
C;Keywords: Intramolecular transferase; isomerase

Query Match 41.1%; Score 44; DB 2; Length 411;  
Best Local Similarity 64.3%; Pred. No. 23;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 EAGNFERISGLKY 15  
||||||| |  
Db 219 EAGNFERTGRRDY 232

RESULT 7  
F85650  
hypothetical protein Z1551 [imported] - Escherichia coli (strain O157:H7)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C;Accession: F85650  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; et al. 2001  
Genome Res. 11:1-16  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: F85650  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-317 <STO>  
A;Cross-references: GB:AE005174; NID:g12514422; PIDN:AAG55666.1; GSPDB:GN00145; UWGP:Z1551  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z1551

Query Match 40.2%; Score 43; DB 2; Length 317;  
Best Local Similarity 50.0%; Pred. No. 25;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 EAGNFERISGLKYQTQID 19  
|:|:|:|:|:|:|  
Db 111 ESGSFEKISSALHTATID 128

RESULT 8  
G72320  
penicillin-binding protein, class 1A - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: G72320  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, J.D.; et al. 1999  
Science 286:2053-2058  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing of *Thermotoga maritima*.  
A;Reference number: A72200; MUID:99287316  
A;Accession: G72320  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-643 <ARN>  
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35967.1; PID:g498142

A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM0886  
C;Superfamily: penicillin-binding protein 1B

Query Match 40.2%; Score 43; DB 2; Length 643;  
Best Local Similarity 37.5%; Pred. No. 57;  
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 5 NFERISGLKYQTQIDQ 20  
::|:|:|:|:|:|:|  
Db 556 SWEKFGDLRYKRLDK 571

RESULT 9  
C84965  
oxoglutarate dehydrogenase (lipoamide) (EC 1.2.4.2) el component [imported] - Buchnera sp.  
C;Species: Buchnera sp.  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C;Accession: C84965  
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.  
Nature 407, 81-86, 2000  
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.  
A;Reference number: A84930; MUID:20445173  
A;Accession: C84965  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-909 <STO>  
A;Cross-references: GB:AP000398; GSPDB:GN00144  
A;Experimental source: strain APS  
C;Genetics:  
A;Gene: suCA; BU302  
C;Superfamily: oxoglutarate dehydrogenase (lipoamide); thiamin pyrophosphate-binding  
C;Keywords: oxidoreductase

Query Match 40.2%; Score 43; DB 2; Length 909;  
Best Local Similarity 70.0%; Pred. No. 84;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 6 FERISGLKY 15  
|:|:|:|:|:|:|  
Db 270 FQKISGDVY 279

RESULT 10  
JC2190  
preprotein translocase secA - Synecococcus sp.  
C;Species: Synecococcus sp.  
C;Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 17-Nov-2000  
C;Accession: JC2190  
R;Nakai, M.; Nohara, T.; Sugita, D.; Endo, T.  
Biochem. Biophys. Res. Commun. 200, 844-851, 1994  
A;Title: Identification and characterization of the sec-A protein homologue in the cytoplasm of *Synechococcus* sp.  
A;Reference number: JC2190; MUID:94235036  
A;Accession: JC2190  
A;Molecule type: DNA  
A;Residues: 1-948 <NAK>  
A;Cross-references: EMBL:X74592; NID:g505309; PIDN:CAA52669.1; PID:g505310  
C;Comment: This protein participates in protein translocation across both the cytoplasmic and the inner membrane. The "nucleotide-binding motif B" and "DEAD motif" features as annotated are adjacent to the identified motif and a third conserved motif is approximately 120-130 residues upstream.  
C;Genetics:  
A;Gene: secA  
C;Superfamily: preprotein translocase secA  
C;Keywords: ATP; homodimer; membrane-associated complex; P-loop; protein transport  
F;106-113/Region: nucleotide-binding motif A (P-loop) #status atypical  
F;209-214/Region: nucleotide-binding motif B  
F;213-216/Region: DEAD motif

Query Match 40.2%; Score 43; DB 2; Length 948;

Best Local Similarity 64.3%; Pred. No. 88;  
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 RISGDLKYTQIDQV 21  
II: III: IIII:  
DB 847 RIAYDLKEAQIDQI 860

## RESULT 11

LEECA  
hemolysin A - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 30-Sep-1988 #sequence\_revision 30-Sep-1988 #text\_change 18-Jun-1999  
C:Accession: A24433; I41280  
R:Feimlee, T.; Pellett, S.; Welch, R.A.  
J. Bacteriol. 163, 94-105, 1985  
A:Title: Nucleotide sequence of an Escherichia coli chromosomal hemolysin.  
A:Reference number: A24433; MUID:85234404.  
A:Accession: A24433  
A:Molecule type: DNA  
A:Residues: 1-1023 <FEL>  
A:Cross-references: GB:M10133; GB:M12863; NID:g146377; PIDN:AAA23975.1; PID:g146379  
A:Experimental source: strain J96, O4 serotype  
R:Stanley, P.; Packman, L.C.; Koronakis, V.; Hughes, C.  
Science 266, 1992-1996, 1994  
A:Title: Fatty acylation of two internal lysine residues required for the toxic activity.  
A:Reference number: A55387; MUID:95099325  
A:Contents: annotation; lysine palmitoylation  
A:Note: lysine modification is performed by the hlyC gene product  
R:Haertlein, M.; Schiessl, S.; Wagner, W.; Rdest, U.; Kreft, J.; Goebel, W.  
J. Cell Biol. 22, 87-97, 1983  
A:Title: Transport of hemolysin by Escherichia coli.  
A:Reference number: I41280  
A:Accession: I41280  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1, 'T', 3, 'V', 5, 'T', 7-44 <RES>  
A:Cross-references: GB:M29173; NID:g146337; PIDN:AAA23957.1; PID:g146338  
C:Genetics:  
C:Function:  
A:Gene: hlyA  
A:Description: attacks blood cell membranes and causes cell lysis  
C:Superfamily: hemolysin A; hemolysin-A homology  
C:Keywords: calcium binding; cytotoxicity; exotoxin; hemolysis; lipoprotein; tandem repeat;  
F:246-791/Domain: hemolysin A homology <HLYA>  
F:723-851/Region: 9-residue repeats (G-G-X-G-[DN]-D-X-[LVIVF]-X)  
F:563,689/Binding site: palmitate (Lys) (covalent) #status experimental

Query Match 40.2%; Score 43; DB 1; Length 1023;

Best Local Similarity 50.0%; Pred. No. 97;  
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 6 PERISGDLKYTQIDQV 21  
II: III: IIII:  
DB 914 FEKSGDISNHQIEQI 929

## RESULT 12

RNA helicase A - human  
C:Species: Homo sapiens (man)  
C:Date: 08-May-1995 #sequence\_revision 12-May-1995 #text\_change 02-Feb-2001  
C:Accession: A47363  
R:Lee, C.G.; Hurwitz, J.  
J. Biol. Chem. 268, 16822-16830, 1993  
A:Title: Human RNA helicase A is homologous to the maleless protein of Drosophila.  
A:Reference number: A47363; MUID:93346440  
A:Accession: A47363  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1279 <LEE>  
A:Cross-references: GB:L13848

A:Note: this sequence follows the published nucleotide sequence and translation and d  
C:Keywords: ATP; nucleotide binding; nucleus; P-loop  
F:411-418/Region: nucleotide-binding motif A (P-loop)  
F:507-512/Region: nucleotide-binding motif B  
F:511-514/Region: DEXH motif

Query Match 40.2%; Score 43; DB 2; Length 1279;  
Best Local Similarity 58.3%; Pred. No. 1.2e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 ERISGDLKYTQI 18  
II: III: IIII:  
DB 192 EKIQGEVKYTVQV 203

## RESULT 13

I46032  
nuclear DNA helicase II - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 02-Feb-2001  
C:Accession: I46032; S49822  
R:Zhang, S.; Maacke, H.; Grosse, F.  
J. Biol. Chem. 270, 16422-16427, 1995  
A:Title: Molecular cloning of the gene encoding nuclear DNA helicase II. A bovine hom  
A:Reference number: I46032; MUID:95332357  
A:Accession: I46032  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1287 <SHA>  
A:Cross-references: EMBL:X82829; NID:g577738; PIDN:CAA58036.1; PID:g577739  
C:Genetics:  
A:Gene: ndh2  
C:Keywords: ATP; nucleotide binding; P-loop  
F:408-415/Region: nucleotide-binding motif A (P-loop)  
F:504-509/Region: nucleotide-binding motif B  
F:508-511/Region: DEXH motif

Query Match 40.2%; Score 43; DB 2; Length 1287;  
Best Local Similarity 58.3%; Pred. No. 1.3e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 ERISGDLKYTQI 18  
II: III: IIII:  
DB 189 EKIQGEVKYTVQV 200

## RESULT 14

JC1276  
dihydroorotate oxidase (EC 1.3.3.1) - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein A314; protein YKL216w  
C:Species: Saccharomyces cerevisiae  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jul-2000  
C:Accession: JC1276; S38059; S44322; S17008  
R:Roy, A.  
Gene 118, 149-150, 1992  
A:Title: Nucleotide sequence of the URA1 gene of Saccharomyces cerevisiae.  
A:Reference number: JC1276; MUID:92380485  
A:Accession: JC1276  
A:Molecule type: DNA  
A:Residues: 1-314 <ROY>  
A:Cross-references: GB:M83295; NID:g171393; PIDN:AAA34566.1; PID:g171394  
R:Alexandraki, D.; Tzermia, M.  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38054  
A:Accession: S38059  
A:Molecule type: DNA  
A:Residues: 1-314 <ALE>  
A:Cross-references: EMBL:228216; NID:g486386; PIDN:CAA82061.1; PID:g486387; GSPDB:GNO  
A:Experimental source: strain S288C  
R:Tzermia, M.; Horaitis, O.; Alexandraki, D.  
Yeast 10, 663-679, 1994

A:Title: The complete sequencing of a 24.6 kb segment of yeast chromosome XI identified e dehydratases, membrane transporters, hydantoinases and the phospholipase A(2)-activator  
A:Reference number: S44319, MUID:95028164  
A:Accession: S44322

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-314 <TZE>

A:Cross-references: EMBL:X75951; NID:g473130; PIDN:CAA53557.1; PID:g473134

C:Genetics:

A:Gene: SGD:URAI; MIPS:YKL216w

A:Cross-references: SGD:S0001699; MIPS:YKL216w

A:Map position: 11L

C:Superfamily: dihydroxotrate oxidase

C:Keywords: flavoprotein; FMN; oxidoreductase; pyrimidine nucleotide biosynthesis

Query Match 39.7%; Score 42.5; DB 1; Length 314;

Best Local Similarity 50.0%; Pred. No. 30;

Matches 11; Conservative 1; Mismatches 3; Indels 7; Gaps 1;

QY 6 FERISGLK-----YTQIDQ 20

||||:|||||

Db 285 FERIEKLDIMEAKGYTSIDQ 306

#### RESULT 15

S75023

sensory transduction histidine kinase sll1905 - *Synechocystis* sp. (strain PCC 6803)

N:Alternate names: protein sll1905

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C:Accession: S75023

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S75023

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1014 <KAN>

A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAAL7885.1; PID:g165296

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: sensory transduction histidine kinase; response regulator homology

C:Keywords: Phosphoprotein

F;20-131/Domain: response regulator homology <RRH1>

F;293-542/Domain: sensor histidine kinase homology <SHK>

F;742-854/Domain: response regulator homology <RRH2>

F;68/Binding site: phosphate (Asp) (covalent) #status predicted

F;790/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match

Best Local Similarity 39.7%; Score 42.5; DB 1; Length 1014;

Matches 8; Conservative 7; Mismatches 3; Indels 1; Gaps 1;

QY 3 AGNFERISGLKVTQIDQV 21

|::||:|::|

Db 997 AASYRVQGLK-TMLDRL 1014

Search completed: July 5, 2001, 11:48:34

Job time: 373 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:51:42 ; Search time 41.8 seconds  
(without alignments)  
17.210 Million cell updates/sec

Title: US-09-462-480-8  
Perfect score: 107  
Sequence: 1 QEAGNFERISGLKYTQIDQV 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	46.5	43.5	3712	1 LMA_DROME	Q00174 drosophila
2	46	43.0	297	1 BMRU_BACSU	P39074 bacillus su
3	46	43.0	547	1 YMX1_CAEEL	P34509 caenorhabdi
4	44	41.1	1	1 DEOB_LACLC	O32808 lactococcus
5	43	40.2	909	1 OD01_BUCAI	P57388 buchnera ap
6	43	40.2	948	1 SECA_SYNP7	O55357 synecococc
7	43	40.2	1023	1 HLX1_ECOLI	P09983 escherichia
8	43	40.2	1270	1 DDX9_HUMAN	Q08211 homo sapien
9	43	40.2	1287	1 DDX9_BOVIN	Q28141 bos taurus
10	43	40.2	1380	1 DDX9_MOUSE	O70133 mus musculu
11	42.5	39.7	314	1 PYRD_YEAST	P28272 saccharomyc
12	42	39.3	462	1 YMA5_YEAST	O04235 saccharomyc
13	42	39.3	831	1 NAH3_RAT	P26433 rattus norv
14	41.5	38.8	77	1 GP46_BPSPI	O48400 bacterioph
15	41.5	38.8	354	1 ARG2_MOUSE	O08691 mus musculu
16	41	38.3	128	1 GYRA_SALTY	P37411 salmonella
17	41	38.3	189	1 LUXP_PHOPO	P25082 photobacter
18	41	38.3	192	1 MORA_HAEIN	P44899 haemophilus
19	41	38.3	392	1 INU2_ARTGO	P19870 arthrobacte
20	41	38.3	875	1 GYRA_ECOLI	P09097 escherichia
21	41	38.3	878	1 GYRA_ERWCA	P41513 erwinnia car
22	41	38.3	1290	1 PIG1_RAT	P10686 rattus norv
23	41	38.3	1511	1 PDRC_YEAST	Q02785 saccharomyc
24	40	37.4	320	1 Y054_MYCPN	P75049 mycoplasma
25	40	37.4	428	1 RFA_KLULA	P09806 kluyveromyc
26	40	37.4	439	1 YM3M_CAEEL	P91343 caenorhabdi
27	40	37.4	697	1 CRAC_DICDI	P35401 dictyosteli
28	40	37.4	735	1 YDD7_SCHPO	Q10432 schizosacch
29	40	37.4	767	1 CC1Q_SCHPO	P01129 schizosacch
30	40	37.4	876	1 GYRA_KLEPN	P14829 klebsiella
31	40	37.4	880	1 GYRA_THIEU	P26221 thermomonos
32	40	37.4	880	1 GYRA_HAEIN	P43700 haemophilus
33	40	37.4	1024	1 HLVA_ECOLI	P08715 escherichia

34	40	37.4	1960	1 MYSN_HUMAN	P35579 homo sapien
35	39.5	36.9	487	1 SYS_CAEEL	Q18678 caenorhabdi
36	39.5	36.9	540	1 MLFS_LACLA	Q48662 lactococcus
37	39	36.4	145	1 YIF1_YEAST	P40185 saccharomyc
38	39	36.4	168	1 ATPF_PROMO	P21904 propionigen
39	39	36.4	186	1 LUXP_PHOLE	Q06877 photobacter
40	39	36.4	217	1 VT11_YEAST	Q04338 saccharomyc
41	39	36.4	218	1 CTR2_VESCR	P00769 vespa crabr
42	39	36.4	308	1 IFRH_SOLTU	P52578 solanum tub
43	39	36.4	342	1 YW12_CAEEL	Q22306 caenorhabdi
44	39	36.4	344	1 PUR5_HAEIN	P43848 haemophilus
45	39	36.4	482	1 GRAA_BACSU	P07868 bacillus su

#### ALIGNMENTS

RESULT 1  
LMA\_DROME  
ID LMA\_DROME STANDARD; PRT; 3712 AA.  
AC Q00174;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE LAMININ ALPHA CHAIN PRECURSOR.  
GN LANA OR LAMA.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93049203; PubMed=1425586;  
RA Kusche-Gulberg M., Garrison K., Mackrell A.J., Fessler L.I.,  
RT Fessler J.H.;  
RT "Laminin A chain: expression during Drosophila development and  
genomic sequence.";  
RL EMBO J. 11:4519-4527(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=94038678; PubMed=8223265;  
RA Henchcliffe C., Garcia-Alonso L., Tang J., Goodman C.S.;  
RT "Genetic analysis of laminin A reveals diverse functions during  
morphogenesis in Drosophila.";  
RL Development 118:325-337(1993).  
RN [3]  
RP SEQUENCE OF 1762-3712 FROM N.A.  
RX MEDLINE=92078147; PubMed=1744083;  
RA Garrison K., Mackrell A.J., Fessler J.H.;  
RT "Drosophila laminin A chain sequence, interspecies comparison, and  
domain structure of a major carboxyl portion.";  
RL J. Biol. Chem. 266:22899-22904(1991).  
CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ  
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
CC -!- FUNCTION: DIVERSE FUNCTIONS DURING MORPHOGENESIS IN DROSOPHILA.  
CC COMPLETE LOSS-OF-FUNCTION MUTATIONS LEAD TO LATE EMBRYONIC  
LETHALITY. CERTAIN PARTIAL LOSS-OF-FUNCTION MUTATIONS GIVE RAISE  
TO ESCAPER ADULTS, WHICH HAVE ROUGH EYES ASSOCIATED WITH CHANGES  
IN CELL FATE AND PATTERN, MISSHAPEN LEGS AND DEFECTS IN WING  
STRUCTURE.  
CC -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT  
MEMBRANES (MAJOR COMPONENT).  
CC -!- TISSUE SPECIFICITY: NEWLY FORMED MESODERM AND LATER PROMINENTLY  
EXPRESSED IN HEMOCYTES, WHICH ALSO SYNTHESIZE COLLAGEN IV.  
CC -!- DEVELOPMENTAL STAGE: DURING MORPHOGENESIS, MOSTLY IN EMBRYO

DEVELOPMENT AT 10-12 HOURS.

-!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

-!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.

-!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

-!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV (DOMAIN IV) IS NOT SIMILAR TO LAMININ DOMAIN IV).

-!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL; M96388; AAC28662.1; -;  
 CC EMBL; L07288; AAC37178.1; -;  
 CC EMBL; M75882; AAC28661.1; -;  
 CC HSSP; P02468; 1TLE.  
 CC FlyBase; FBgn0002526; LANA.  
 CC InterPro; IPR000034; -;  
 CC InterPro; IPR000561; -;  
 CC InterPro; IPR001791; -;  
 CC InterPro; IPR001886; -;  
 CC InterPro; IPR002049; -;  
 CC Pfam; PF00052; laminin\_B; 1.  
 CC Pfam; PF00053; laminin\_EGF; 20.  
 CC Pfam; PF00054; laminin\_G; 5.  
 CC Pfam; PF00055; laminin\_Nterm; 1.  
 CC PRINTS; PR00011; EGF\_LAMININ.  
 CC PROSITE; PS00022; EGF\_1; 17.  
 CC PROSITE; PS01186; EGF\_2; 5.  
 CC PROSITE; PS01248; LAMININ\_TYPE\_EGF; 19.  
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 3712 LAMININ ALPHA CHAIN.  
 FT DOMAIN 25 272 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 273 815 10.5 X LAMININ EGF-LIKE REPEATS (DOMAIN V).  
 FT DOMAIN 273 332 LAMININ EGF-LIKE 1.  
 FT DOMAIN 333 402 LAMININ EGF-LIKE 2.  
 FT DOMAIN 403 447 LAMININ EGF-LIKE 3.  
 FT DOMAIN 448 494 LAMININ EGF-LIKE 4.  
 FT DOMAIN 495 540 LAMININ EGF-LIKE 5.  
 FT DOMAIN 541 586 LAMININ EGF-LIKE 6.  
 FT DOMAIN 587 631 LAMININ EGF-LIKE 7.  
 FT DOMAIN 632 676 LAMININ EGF-LIKE 8.  
 FT DOMAIN 677 731 LAMININ EGF-LIKE 9.  
 FT DOMAIN 732 784 LAMININ EGF-LIKE 10.  
 FT DOMAIN 785 815 LAMININ EGF-LIKE 11 (INCOMPLETE).  
 FT DOMAIN 816 1374 DOMAIN IV'.  
 FT DOMAIN 1375 1574 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN III B).  
 FT DOMAIN 1375 1420 LAMININ EGF-LIKE 12.  
 FT DOMAIN 1421 1465 LAMININ EGF-LIKE 13.  
 FT DOMAIN 1466 1513 LAMININ EGF-LIKE 14.  
 FT DOMAIN 1514 1564 LAMININ EGF-LIKE 15.  
 FT DOMAIN 1565 1574 LAMININ EGF-LIKE 16 (N-TERMINAL).  
 FT DOMAIN 1575 1775 LAMININ DOMAIN IV (DOMAIN IV).  
 FT DOMAIN 1776 2111 6.5 X LAMININ EGF-LIKE REPEATS (DOMAIN III A).  
 FT DOMAIN 1776 1808 LAMININ EGF-LIKE 16 (C-TERMINAL).  
 FT DOMAIN 1809 1858 LAMININ EGF-LIKE 17.  
 FT DOMAIN 1859 1916 LAMININ EGF-LIKE 18.  
 FT DOMAIN 1917 1969 LAMININ EGF-LIKE 19.  
 FT DOMAIN 1970 2016 LAMININ EGF-LIKE 20.  
 FT DOMAIN 2017 2063 LAMININ EGF-LIKE 21.  
 FT DOMAIN 2064 2111 LAMININ EGF-LIKE 22.  
 FT DOMAIN 2112 2697 DOMAIN II AND I.

FT DOMAIN 2698 3712 5 X LAMININ G-LIKE REPEATS (DOMAIN G).  
 FT DOMAIN 2698 2862 LAMININ G-LIKE 1.  
 FT DOMAIN 2863 3048 LAMININ G-LIKE 2.  
 FT DOMAIN 3049 3223 LAMININ G-LIKE 3.  
 FT DOMAIN 3224 3296 POLY-THR.  
 FT DOMAIN 3334 3528 LAMININ G-LIKE 4.  
 FT DOMAIN 3529 3712 LAMININ G-LIKE 5.  
 FT DOMAIN 2178 2249 COILED COIL (POTENTIAL).  
 FT DOMAIN 2301 2321 COILED COIL (POTENTIAL).  
 FT DOMAIN 2376 2450 COILED COIL (POTENTIAL).  
 FT DOMAIN 2541 2676 COILED COIL (POTENTIAL).  
 FT DISULFID 273 282 BY SIMILARITY.  
 FT DISULFID 275 296 BY SIMILARITY.  
 FT DISULFID 298 307 BY SIMILARITY.  
 FT DISULFID 310 330 BY SIMILARITY.  
 FT DISULFID 333 342 BY SIMILARITY.  
 FT DISULFID 335 367 BY SIMILARITY.  
 FT DISULFID 370 379 BY SIMILARITY.  
 FT DISULFID 382 400 BY SIMILARITY.  
 FT DISULFID 403 414 BY SIMILARITY.  
 FT DISULFID 405 421 BY SIMILARITY.  
 FT DISULFID 423 432 BY SIMILARITY.  
 FT DISULFID 435 445 BY SIMILARITY.  
 FT DISULFID 448 460 BY SIMILARITY.  
 FT DISULFID 450 468 BY SIMILARITY.  
 FT DISULFID 470 479 BY SIMILARITY.  
 FT DISULFID 482 492 BY SIMILARITY.  
 FT DISULFID 495 507 BY SIMILARITY.  
 FT DISULFID 497 514 BY SIMILARITY.  
 FT DISULFID 516 525 BY SIMILARITY.  
 FT DISULFID 528 538 BY SIMILARITY.  
 FT DISULFID 541 553 BY SIMILARITY.  
 FT DISULFID 543 560 BY SIMILARITY.  
 FT DISULFID 562 571 BY SIMILARITY.  
 FT DISULFID 574 584 BY SIMILARITY.  
 FT DISULFID 587 599 BY SIMILARITY.  
 FT DISULFID 589 605 BY SIMILARITY.  
 FT DISULFID 607 616 BY SIMILARITY.  
 FT DISULFID 619 629 BY SIMILARITY.  
 FT DISULFID 632 644 BY SIMILARITY.  
 FT DISULFID 634 650 BY SIMILARITY.  
 FT DISULFID 652 661 BY SIMILARITY.  
 FT DISULFID 664 674 BY SIMILARITY.  
 FT DISULFID 677 691 BY SIMILARITY.  
 FT DISULFID 679 700 BY SIMILARITY.  
 FT DISULFID 702 711 BY SIMILARITY.  
 FT DISULFID 714 729 BY SIMILARITY.  
 FT DISULFID 732 746 BY SIMILARITY.  
 FT DISULFID 734 753 BY SIMILARITY.  
 FT DISULFID 755 764 BY SIMILARITY.  
 FT DISULFID 767 782 BY SIMILARITY.  
 FT DISULFID 1375 1387 BY SIMILARITY.  
 FT DISULFID 1377 1394 BY SIMILARITY.  
 FT DISULFID 1396 1405 BY SIMILARITY.  
 FT DISULFID 1408 1418 BY SIMILARITY.  
 FT DISULFID 1421 1429 BY SIMILARITY.  
 FT DISULFID 1432 1436 BY SIMILARITY.  
 FT DISULFID 1438 1447 BY SIMILARITY.  
 FT DISULFID 1450 1463 BY SIMILARITY.  
 FT DISULFID 1466 1480 BY SIMILARITY.  
 FT DISULFID 1468 1487 BY SIMILARITY.  
 FT DISULFID 1489 1498 BY SIMILARITY.  
 FT DISULFID 1501 1511 BY SIMILARITY.  
 FT DISULFID 1514 1526 BY SIMILARITY.  
 FT DISULFID 1516 1533 BY SIMILARITY.  
 FT DISULFID 1535 1544 BY SIMILARITY.  
 FT DISULFID 1547 1562 BY SIMILARITY.  
 FT DISULFID 1859 1874 BY SIMILARITY.  
 FT DISULFID 1861 1885 BY SIMILARITY.  
 FT DISULFID 1887 1896 BY SIMILARITY.  
 FT DISULFID 1899 1914 BY SIMILARITY.  
 FT DISULFID 1917 1931 BY SIMILARITY.  
 FT DISULFID 1919 1938 BY SIMILARITY.

RESULT	3
YMX1_CAEEL	
ID	YMX1_CAEEL
AC	P34509;
DT	01-FEB-1994 (Rel. 28, Created)
DT	01-FEB-1994 (Rel. 28, Last sequence update)
DT	01-NOV-1995 (Rel. 32, Last annotation update)
DE	PUTATIVE SERINE/THREONINE-PROTEIN KINASE K06H7.1 IN CHROMOSOME III

RESULT	4	
DEOB_LACLC		
ID	DEOB_LACLC	STANDARD; PRT; 411 AA.
AC	O32808;	
DT	15-JUL-1998	(Rel. 36, Created)
DT	15-JUL-1998	(Rel. 36, Last sequence update)
DT	15-DEC-1998	(Rel. 37, Last annotation update)
DE	PHOSPHOPENTOMUTASE (EC 5.4.2.7)	(PHOSPHOEXYRIBOMUTASE).
GN	DEOB.	
OS	Lactococcus lactis (subsp. cremoris)	(Streptococcus cremoris).



```
QY 8 RISGLKTYQIDQV 21
   ||: ||| |||:
Db 847 RIAYDLKEAQIDQI 860

RESULT 7
HLV1_ECOLI STANDARD; PRT; 1023 AA.
AC P09983;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMOLYSIN, CHROMOSOMAL.
GN HLYA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J96 / SEROTYPE O4;
RX MEDLINE=85234404; PubMed=3891743;
RA Felmler T., Pellett S., Welch R.A.;
RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";
RL J. Bacteriol. 163:94-105(1985).
RN [2]
RP SEQUENCE OF 1-44 FROM N.A.
RC STRAIN=2001;
RX MEDLINE=85258115; PubMed=3894051;
RA Nicaud J.-M., Mackman N., Gray L., Holland I.B.;
RT "Characterisation of HlyC and mechanism of activation and secretion
of haemolysin from E. coli 2001.";
RL FEBS Lett. 187:339-344(1985).
CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
DEFINED.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
ACTIVITY.
CC -1- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE
INVOLVED IN PORE FORMATION BY THE CYTOTOXIN.
CC -1- PTM: PALMITOYLATED BY HLYC. THE TOXIN ONLY BECOMES ACTIVE WHEN
MODIFIED.
CC -1- DISEASE: THE HEMOLYSIN OF E. COLI IS PRODUCED PREDOMINANTLY BY
STRAINS CAUSING EXTRAINTestinal INFECTIONS, SUCH AS THOSE OF THE
URINARY TRACT.
CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M10133; AAA23975.1; -
DR EMBL; X02768; CAA26546.1; -
DR PIR; A24433; LEECA.
DR InterPro; IPR001343; -
DR Pfam; PF00353; hemolysinCbind; 2.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00330; HEMOLYSIN_CALCIUM; 4.
KW Hemolysin; Toxin; Cytolysin; Cytotoxin; Repeat; Calcium;
KW Transmembrane; Lipoprotein; Palmitate.
FT TRANSMEM 237 259 POTENTIAL.
FT TRANSMEM 267 326 POTENTIAL.
FT TRANSMEM 364 410 POTENTIAL.
FT DOMAIN 723 869 16 X REPEATS, GLY-RICH.
FT REPEAT 732 737 2.
FT REPEAT 741 746 3.
```

```
FT REPEAT 750 755 4.
FT REPEAT 759 764 5.
FT REPEAT 768 773 6.
FT REPEAT 777 782 7.
FT REPEAT 786 791 8.
FT REPEAT 795 800 9.
FT REPEAT 806 812 10.
FT REPEAT 816 821 11.
FT REPEAT 825 830 12.
FT REPEAT 834 839 13.
FT REPEAT 843 848 14.
FT REPEAT 855 860 15.
FT REPEAT 864 869 16.
FT LIPID 563 563 PALMITATE (BY SIMILARITY).
FT LIPID 689 689 PALMITATE (BY SIMILARITY).
FT VARIANT 6 A -> T (IN STRAIN 2001).
SQ SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64;

Query Match 40.2%; Score 43; DB 1; Length 1023;
Best Local Similarity 50.0%; Pred No. 28;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 6 FERISGLKTYQIDQV 21
   ||: ||| |||:
Db 914 FEKESGDISHNQIEQI 929

RESULT 8
DDX9_HUMAN STANDARD; PRT; 1270 AA.
AC Q08211; Q99556;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ATP-DEPENDENT RNA HELICASE A (NUCLEAR DNA HELICASE II) (NDH II)
DE (DEAD-BOX PROTEIN 9).
GN DDX9 OR NDH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93346440; PubMed=8344961;
RA Lee C.-G., Hurwitz J.;
RT "Human RNA helicase A is homologous to the maleless protein of
Drosophila.";
RL J. Biol. Chem. 268:16822-16830(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97269062; PubMed=9111062;
RA Zhang S., Grosse F.;
RT "Domain structure of human nuclear DNA helicase II (RNA helicase A).";
RL J. Biol. Chem. 272:11487-11494(1997).
CC -1- FUNCTION: UNWINDS DOUBLE-STRANDED DNA AND RNA IN A 3' TO 5'
DIRECTION. GENERATES MULTIPLE SECONDARY STRUCTURES THAT INFLUENCE
RNA-BINDING PROTEINS. MAY PLAY A ROLE IN X-LINKED GENE EXPRESSION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE. DEAD
SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; LI3848; AAB48855.1; -
DR EMBL; Y10658; CAA71668.1; -
```

MIM: 603115; --  
 DR InterPro: IPR001159; --  
 DR InterPro: IPR001650; --  
 DR InterPro: IPR002464; --  
 DR Pfam: PF00035; dsrm; 2.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.  
 DR PROSITE; PS50137; DS\_RBD; 2.  
 KW Helicase; RNA-binding; DNA-binding; Repeat; Nuclear protein;  
 KW ATP-binding.  
 FT DOMAIN 3 71 DDBM 1.  
 FT NP\_BIND 180 252 DDBM 2.  
 FT SITE 411 418 ATP (BY SIMILARITY).  
 FT SITE 511 515 DEIH BOX.  
 FT DOMAIN 586 595 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 1170 1270 ARG/GLY/SER/TYR-RICH.  
 FT CONFLICT 20 20 S -> T (IN REF. 1).  
 FT CONFLICT 108 109 TM -> HH (IN REF. 1).  
 FT CONFLICT 114 116 PPH -> LHI (IN REF. 1).  
 FT CONFLICT 186 186 N -> I (IN REF. 1).  
 FT CONFLICT 260 260 S -> T (IN REF. 1).  
 FT CONFLICT 478 478 I -> V (IN REF. 1).  
 FT CONFLICT 521 521 D -> S (IN REF. 1).  
 FT CONFLICT 541 541 L -> F (IN REF. 1).  
 FT CONFLICT 560 565 IIEVYG -> SLKLM (IN REF. 1).  
 FT CONFLICT 894 894 I -> V (IN REF. 1).  
 FT CONFLICT 899 899 R -> Q (IN REF. 1).  
 FT CONFLICT 1037 1037 K -> N (IN REF. 1).  
 FT CONFLICT 1140 1140 R -> E (IN REF. 1).  
 FT CONFLICT 1204 1211 NSFRAGYG -> TPSSGRC (IN REF. 1).  
 FT CONFLICT 1261 1270 FGGRGGGGY -> LDIEEVAAILKGYSSVCRCQ (IN REF. 1).  
 SQ SEQUENCE 1270 AA; 140876 MW; 3DEBB9ED3D6C9B2D CRC64;

Query Match 40.2%; Score 43; DB 1; Length 1270;  
 Best Local Similarity 58.3%; Pred. No. 36;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 ERISGLKYTOI 18  
 I:| |:| | | | |  
 DB 192 EKIQGEYKYTOV 203

RESULT 9  
 DDX9\_BOVIN  
 ID DDX9\_BOVIN STANDARD; PRT; 1287 AA.  
 AC Q28141; 1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ATP-DEPENDENT RNA HELICASE A (NUCLEAR DNA HELICASE II) (NDH II)  
 DE (DEAD-BOX PROTEIN 9).  
 GN DDX9 OR NDH2.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Thymus;  
 RX MEDLINE=95332357; PubMed=7608213;  
 RA Zhang S., Maacke H., Grosse F.;  
 RT "Molecular cloning of the gene encoding nuclear DNA helicase A and Drosophila Mle protein.";  
 RL J. Biol. Chem. 270:16422-16427(1995).  
 CC -1- FUNCTION: UNWINDS DOUBLE-STRANDED DNA AND RNA IN A 3' TO 5' DIRECTION. GENERATES MULTIPLE SECONDARY STRUCTURES THAT INFLUENCE RNA-BINDING PROTEINS. MAY PLAY A ROLE IN X-LINKED GENE EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE. DEAH  
 CC -1- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE. DEAH

CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 DDBM (DOUBLE-STRANDED RNA-BINDING) DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X82829; CAA58036.1; --  
 DR InterPro: IPR001159; --  
 DR InterPro: IPR001650; --  
 DR InterPro: IPR002464; --  
 DR Pfam: PF00035; dsrm; 2.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.  
 DR PROSITE; PS50137; DS\_RBD; 2.  
 KW Helicase; RNA-binding; DNA-binding; Repeat; Nuclear protein;  
 KW ATP-binding.  
 FT DOMAIN 3 71 DDBM 1.  
 FT NP\_BIND 177 249 DDBM 2.  
 FT SITE 508 511 ATP (BY SIMILARITY).  
 FT SITE 583 592 DEIH BOX.  
 FT DOMAIN 1167 1287 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 1287 AA; 141943 MW; DC908095AB683ED4 CRC64;  
 SQ SEQUENCE 1287 AA; 141943 MW; DC908095AB683ED4 CRC64;

Query Match 40.2%; Score 43; DB 1; Length 1287;  
 Best Local Similarity 58.3%; Pred. No. 37;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 ERISGLKYTOI 18  
 I:| |:| | | | |  
 DB 189 EKIQGEYKYTOV 200

RESULT 10  
 DDX9\_MOUSE  
 ID DDX9\_MOUSE STANDARD; PRT; 1380 AA.  
 AC O70133; O35931; O54703;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ATP-DEPENDENT RNA HELICASE A (NUCLEAR DNA HELICASE II) (NDH II)  
 DE (DEAD-BOX PROTEIN 9) (MHEL-5).  
 GN DDX9.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=98149984; PubMed=9480750;  
 RA Lee C.-G., Eki T., Okumura K., da Costa Soares V., Hurwitz J.;  
 RT "Molecular analysis of the cDNA and genomic DNA encoding mouse RNA helicase A.";  
 RL Genomics 47:365-371(1998).  
 [2]  
 RN SEQUENCE OF 386-919 FROM N.A.  
 RC STRAIN=C57BL/6;  
 RA Kisielow P., Miazek A.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: UNWINDS DOUBLE-STRANDED DNA AND RNA IN A 3' TO 5' DIRECTION. GENERATES MULTIPLE SECONDARY STRUCTURES THAT INFLUENCE RNA-BINDING PROTEINS. MAY PLAY A ROLE IN X-LINKED GENE EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE. DEAH  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 DDBM (DOUBLE-STRANDED RNA-BINDING) DOMAINS.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U91922; AAC05725.1; -;  
DR EMBL; AF023530; AAC05301.1; -;  
DR EMBL; G92080; AAB72087.1; -;  
DR MGI; 108177; Ddx9.  
DR InterPro; IPR001159; -;  
DR InterPro; IPR001650; -;  
DR InterPro; IPR002464; -;  
DR Pfam; PF00035; dsrm; 2;  
DR Pfam; PF00271; helicase\_C; 1;  
DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1;  
DR PROSITE; PS0137; DS\_RBD; 2;  
KW Helicase; RNA-binding; DNA-binding; Repeat; Nuclear protein;  
ATP-binding.  
FT DOMAIN 3 71 DBRM 1.  
FT DOMAIN 182 254 DBRM 2.  
FT NP\_BIND 413 420 ATP (POTENTIAL).  
FT SITE 513 516 DEAH BOX.  
FT DOMAIN 588 597 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 1171 1380 ARG/GLY/SER/TYR-RICH.  
FT CONFLICT 46 46 R -> A (IN AAC05301).  
FT CONFLICT 674 674 N -> M (IN REF. 2).  
FT CONFLICT 748 748 T -> I (IN REF. 2).  
FT CONFLICT 831 831 I -> V (IN REF. 2).  
SQ SEQUENCE 1380 AA; 149582 MW; 0F2375C291D2FE1 CRC64;

Query Match 40.2%; Score 43; DB 1; Length 1380;  
Best Local Similarity 58.3%; Pred. No. 40;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 ERISGDLKYYTQI 18  
|:|:|:|:|:|:  
Db 194 EKIQGEYKYYTQV 205

RESULT 11  
PYRD\_YEAST STANDARD; PRT; 314 AA.  
AC P28272;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE DIHYDROOROTATE DEHYDROGENASE (EC 1.3.3.1) (DIHYDROOROTATE OXIDASE) (DHODHASE).  
GN URAL OR YKL216W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 28383 / FL100;  
RX MEDLINE=92380485; PubMed=1511880;  
RA ROY A.;  
RT "Nucleotide sequence of the URAL gene of Saccharomyces cerevisiae.";  
RL Gene 118:149-150(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=95028164; PubMed=7941750;  
RX MEDLINE=95028164; PubMed=7941750;  
RA Tzermia M., Horaitis O., Alexandraki D.;  
RT "The complete sequencing of a 24.6 kb segment of yeast chromosome XI identified the known loci URAL1, SAC1 and TRP3, and revealed 6 new open reading frames including homologues to the threonine

RT dehydratases, membrane transporters, hydantoinsases and the  
RL phospholipase A2-activating protein.";  
Yeast 10:663-679(1994). L-DIHYDROOROTATE + O(2) - OROATE + H(2)O(2).  
CC -1- CATALYTIC ACTIVITY: L-DIHYDROOROTATE + O(2) - OROATE + H(2)O(2).  
CC -1- COFACTOR: FAD.  
CC -1- PATHWAY: FOURTH STEP IN PYRIMIDINE BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X59371; CAA42014.1; -;  
DR EMBL; X75951; CAA53557.1; -;  
DR EMBL; M83295; AAA34566.1; -;  
DR EMBL; Z28216; CAA82061.1; -;  
DR PIR; JC1276; JC1276.  
DR PIR; S44322; S44322.  
DR HSP; P54321; 1DOR.  
DR SGD; S0001899; URAL.  
DR InterPro; IPR001295; -;  
DR Pfam; PF01180; DHODHase; 1.  
DR PROSITE; PS00911; DHODHASE\_1; 1.  
DR PROSITE; PS00912; DHODHASE\_2; 1.  
KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FAD.  
FT NP\_BIND 246 254 FAD (NAD PART) (POTENTIAL).  
SQ SEQUENCE 314 AA; 34801 MW; 0F1FF9BDA7F8D68E CRC64;

Query Match 39.7%; Score 42.5; DB 1; Length 314;  
Best Local Similarity 50.0%; Pred. No. 8.9;  
Matches 11; Conservative 1; Mismatches 3; Indels 7; Gaps 1;

QY 6 FERISGDLK-----YTQIDQ 20  
|:|:|:|:|:|:  
Db 285 FERIEKLKDIMEAKGYTSDQ 306

RESULT 12  
YMA5\_YEAST STANDARD; PRT; 462 AA.  
ID YMA5\_YEAST  
AC Q04235;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE HYPOTHETICAL 52.7 KDA PROTEIN IN PDR4-GLO1 INTERGENIC REGION.  
GN YML005W OR YM9571.14.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Gentles S., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO S.POMBE SPAC4G8.06C.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; Z49810; CAA89947.1; -;  
DR SGD; S0004454; YML005W.  
KW Hypothetical protein.





```

CC
ENBL; U90886; AAC22548.1; -.
DR ENBL; AFQ32466; AAB86959.1; -.
DR ENBL; AF045965; AAC78460.1; -.
DR ENBL; AF044680; AAC78460.1; JOINED.
DR ENBL; AF045959; AAC78460.1; JOINED.
DR ENBL; AF045960; AAC78460.1; JOINED.
DR ENBL; AF045961; AAC78460.1; JOINED.
DR ENBL; AF045962; AAC78460.1; JOINED.
DR ENBL; AF045963; AAC78460.1; JOINED.
DR ENBL; AF045964; AAC78460.1; JOINED.
DR HSP; P07824; IRLA.
DR MGD; MGI:1330806; Arg2.
DR InterPro; IPR000287; -.
DR Pfam; PF00491; arginase; 1.
DR PRINTS; PR00116; ARGINASE.
DR PROSITE; PS00147; ARGINASE_1; 1.

```



```

RESULT      1
ID          O69739
AC          PRELIMINARY; PRT; 100 AA.
DT          O69739;
DD          01-AUG-1998 (TrEMBLrel. 07, Created)
DE          01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE          01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE          HYPOTHETICAL 10.8 KDA PROTEIN.
GN          LHP OR RV3874 OR MTV027.09.
GS          Mycobacterium tuberculosis.
OS          Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC          Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
[1]
RNP         SEQUENCE FROM N.A.
RC          STRAIN=H37RV.
RRX         MEDLINE=98295987; PubMed=9634230;
RRY         Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA          Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA          Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA          Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA          Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA          Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA          Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA          Taylor K., Whitehead S., Barrall B.G.;
RA          "Deciphering the biology of Mycobacterium tuberculosis from the
RFT         complete genome sequence.";
RFT         Nature 393:537-544(1998).
RN          [2]
RNP         SEQUENCE FROM N.A.
RC          STRAIN=H37RV;
RRX         Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;
RRY         "Promoter analysis of the M. tuberculosis orfIC gene encoding the
RA          early secreted antigenic target 6 kDa (ESAT-6).";
RA          Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RD          EMBL; AL022120; CAA17966.1; -
DR          EMBL; AF004671; AAC83445.1; -.
DR          Tuberculinist; Kv3874; -.
DR          Hypothetical protein.
KW          SEQUENCE 100 AA; 10794 MW; 285F4FC96F55D194 CRC64;

```

Query Match 83.6%; Score 89.5; DB 2; Length 100;  
 Best Local Similarity 95.2%; Pred. No. 4.5e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 QPAGNFERISGLKTYTQIDQV 21  
 Db 13 QPAGNFERISGLK-TQIDQV 32

## RESULT

Q9XWB8 PRELIMINARY; PRT; 296 AA.

AC Q9XWB8; (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DE Y75B12B.1 PROTEIN.

GN Y75B12B.1

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA White S.;

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,

RA Smaiden N., Smith A., Sonnenhammer E., Staden K., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkins-Sproat J., Wohlman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans";

RL Nature 368:32-38(1994).

DR EMBL; AL032683; CAA21764.1;

SQ SEQUENCE 296 AA; 34018 MW; E39256933B926E91 CRC64;

Query Match 47.7%; Score 51; DB 5; Length 296;

Best Local Similarity 50.0%; Pred. No. 2.7;

Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 QPAGNFERISGLKTYQI 18

Db 187 EETSQLRVSGSKYLQI 204

## RESULT

Q9VRW0 PRELIMINARY; PRT; 3712 AA.

AC Q9VRW0;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE LANA PROTEIN.

GN LANA OR CG10236.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003563; AAF50672.2;  
 DR HSSP; P02468; ITLE.  
 DR FlyBase; FBgn0002526; LANA.  
 DR InterPro; IPR000034; -  
 DR InterPro; IPR000345; -  
 DR InterPro; IPR000504; -  
 DR InterPro; IPR000508; -  
 DR InterPro; IPR000561; -  
 DR InterPro; IPR001230; -  
 DR InterPro; IPR001791; -  
 DR InterPro; IPR001886; -  
 DR InterPro; IPR002049; -  
 DR Pfam; PF00052; laminin\_B; 1.  
 DR Pfam; PF00053; laminin\_EGF; 20.  
 DR Pfam; PF00054; laminin\_G; 5.  
 DR Pfam; PF00055; laminin\_Nterm; 1.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_17.  
 DR PROSITE; PS01186; EGF\_2; 5.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 19.  
 DR PROSITE; PS00294; PRENYLATION; UNKNOWN\_1.  
 DR PROSITE; PS00030; RNP\_1; UNKNOWN\_1.  
 DR PROSITE; PS00761; SPASE\_1\_3; UNKNOWN\_1.  
 KW EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.  
 SQ SEQUENCE 3712 AA; 411126 MW; 2B626D435E06B84A CRC64;

Query Match 43.5%; Score 46.5; DB 5; Length 3712;

Best Local Similarity 55.0%; Pred. No. 2.7e+02;

Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

Qy 1 QPAGNFERISGLKTYTQIDQ 20

Db 2366 KDAGNF-LINGDLTLNQINQ 2384

```

RESULT 4
Q9W766 PRELIMINARY; PRT; 108 AA.
AC Q9W766;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MATRIX METALLOPROTEINASE MMP-9 (FRAGMENT).
OS Ambystoma mexicanum (Axolotl).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Ambystomatidae;
OC Ambystoma.
OX NCBI_TaxID=8296;
RN [1]
SEQUENCE FROM N.A.
RA Yang E.V., Gardiner D.M., Carlson M.R.J., Nugas C.A., Bryant S.V.;
RT "Expression of matrix metalloproteinase genes during axolotl limb
RT regeneration."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF150055; AAD40666.1; -
DR HSSP; P03956; ICGL.
DR InterPro; IPR000130; -
DR InterPro; IPR001818; -
DR Pfam; PF00413; Peptidase_M10; 1.
DR PRINTS; PR00138; MATRIXIN.
DR SMART; SM00235; ZmMc; 1.
FT NON_TER 1
FT NON_TER 108
SQ SEQUENCE 108 AA; 12105 MW; 373757F60B73303C CRC64;

Query Match 43.0%; Score 46; DB 13; Length 108;
Best Local Similarity 44.4%; Pred. No. 5.5;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EAGNFERISGLKYQTID 19
: |||: |||: |
Db 3 DVGNFQTFDGLKWDHND 20

RESULT 5
Q9REB6 PRELIMINARY; PRT; 653 AA.
AC Q9REB6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE MOBB PROTEIN.
GN MOBB.
OS Enterobacter cloacae.
OG Plasmid CloDF13.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=550;
RN [1]
SEQUENCE FROM N.A.
RA Nunez B., Avila P., de la Cruz F.;
RT "Mobilization functions of CloDF13."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ224861; CAB62409.1; -
KW Plasmid.
SQ SEQUENCE 653 AA; 72897 MW; D69F5922FEE8A79F CRC64;

Query Match 43.0%; Score 46; DB 2; Length 653;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QEAGNFERISGD 12
: ||||| ||||
Db 560 REAGNVEHISGD 571

```

```

RESULT 6
Q98856 PRELIMINARY; PRT; 679 AA.
AC Q98856;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE 92 KDA TYPE IV COLLAGENASE PRECURSOR (EC 3.4.24.35) (92 KDA
DE GELATINASE) (MATRIX METALLOPROTEINASE-9) (MMP-9) (GELATINASE B).
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Cynops.
OX NCBI_TaxID=8330;
RN [1]
SEQUENCE FROM N.A.
RA Tissue-REGENERATING LIMB BLASTOMA;
RX MEDLINE=96270627; PubMed=8692902;
RA Miyazaki K., Uchiyama K., Imokawa Y., Yoshizato K.;
RT "Cloning and characterization of cDNAs for matrix metalloproteinases
RT of regenerating newt limbs."
RL Proc. Natl. Acad. Sci. U.S.A. 93:6819-6824(1996).
CC -!- FUNCTION: COULD PLAY A ROLE IN BONE OSTEOCLASTIC RESORPTION. MAY
CC BE INVOLVED IN TISSUE REMODELLING OF THE REGENERATING LIMB.
CC -!- CATALYTIC ACTIVITY: CLEAVAGE OF GELATIN TYPES I AND V AND COLLAGEN
CC TYPES IV AND V.
CC -!- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC -!- SIMILARITY: THE CENTRAL DOMAIN SHOWS SIMILARITY TO THE TYPE-II
CC MOTIF OF COLLAGEN-BINDING DOMAIN OF FIBRONECTIN. THIS DOMAIN BINDS
CC GELATIN.
CC -!- SIMILARITY: CONTAINS 1 HEMOPLEXIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE II-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
CC METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.
DR EMBL; D82052; BAAL1523.1; -
DR HSSP; P08254; IUSN.
DR MEROPS; M10.004; -
DR InterPro; IPR000130; -
DR InterPro; IPR000562; -
DR InterPro; IPR000585; -
DR InterPro; IPR001818; -
DR Pfam; PF00040; fn2; 3.
DR Pfam; PF00045; hemopexin; 4.
DR PRINTS; PR00013; ENTYPEII.
DR PRINTS; PR00138; MATRIXIN.
DR ProDom; PD000995; -; 3.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
DR PROSITE; PS00023; FIBRONECTIN_2; 3.
DR PROSITE; PS00024; HEMOPLEXIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SMART; SM00120; HA; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Zinc; Zymogen; Calcium;
KW Collagen degradation; Extracellular matrix; Signal; Polymorphism.
FT SIGNAL 1 19
FT PROPEP 20 107
FT CHAIN 108 679
FT DOMAIN 98 105
FT DOMAIN 224 281
FT DOMAIN 282 339
FT DOMAIN 340 397
FT DOMAIN 486 679
FT METAL 401 401
FT ACT_SITE 402 402
FT METAL 405 405
FT METAL 411 411
FT DISULFID 491 676
FT CARBOHYD 82 82
SQ SEQUENCE 679 AA; 75581 MW; 573CDD9DBF85524 CRC64;

Query Match 43.0%; Score 46; DB 13; Length 679;

```

Best Local Similarity 44.4%; Pred. No. 46;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 EAGNERISGLKYQID 19  
O9F799 : ||||: ||||: |  
Db 104 DVGNFQTFDGLKWDND 121

RESULT 7  
ID Q9F799 PRELIMINARY; PRT; 464 AA.  
AC Q9F799;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE PUTATIVE UNDECAPENYLPHOSPHATE GLUCOSEPHOSPHOTRANSFERASE.  
GN WCAJ.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Salmonella.

OX NCBI\_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=LT2;

RX MEDLINE=20461159; PubMed=11004393;

RA Stevenson G., Lan R., Reeves P.R.;

RT "The colanic acid gene cluster of salmonella enterica has a complex

history.";

RL FEMS Microbiol. Lett. 191:11-16(2000).

DR EMBL; AF285085; AAG24818.1; -.

KW Transferase.

SQ SEQUENCE 464 AA; 52548 MW; B36027A2231957DD CRC64;

Query Match 41.1%; Score 44; DB 2; Length 464;  
Best Local Similarity 47.4%; Pred. No. 63;  
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGNFERISGLKYQIDQV 21  
||||: | : | |  
Db 190 AGNFEQLIDDAKASRIHV 208

RESULT 8  
ID Q9FA27 PRELIMINARY; PRT; 610 AA.  
AC Q9FA27;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE CWP66.  
GN CWP66.  
OS Clostridium difficile.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
CC Clostridium.  
OX NCBI\_TaxID=1496;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=79-685;

RA Waligora A.J., Bourlioux P., Karjalainen T.;

RT "Cloning of a gene encoding an autolysin-like protein from Clostridium

difficile.";

RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF093106; AAG13990.1; -.

SQ SEQUENCE 610 AA; 66323 MW; C4720E5E0AE28F7B CRC64;

Query Match 41.1%; Score 44; DB 2; Length 610;  
Best Local Similarity 72.7%; Pred. No. 86;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 NFERISGLKY 15  
||||||: |

Db 128 NFERISGNSRY 138

RESULT 9  
ID Q9F9G7 PRELIMINARY; PRT; 610 AA.  
AC Q9F9G7;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE CWP66.  
GN CWP66.

OS Clostridium difficile.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

CC Clostridium.

OX NCBI\_TaxID=1496;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C253;

RA Waligora A.J., Hennequin C., Mullany P., Collignon A., Bourlioux P.;

RA Karjalainen T.;

RT "Cloning, characterization and expression of a gene encoding a surface

protein of C. difficile with adhesive properties.";

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF194870; AAG28425.1; -.

SQ SEQUENCE 610 AA; 66765 MW; 8ADEFC23D1AA1140 CRC64;

Query Match 41.1%; Score 44; DB 2; Length 610;  
Best Local Similarity 72.7%; Pred. No. 86;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 NFERISGLKY 15  
||||||: |

Db 128 NFERISGNSRY 138

RESULT 10  
ID Q9EYK0 PRELIMINARY; PRT; 306 AA.  
AC Q9EYK0;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE PUTATIVE TRANSPOSASE PRCB.  
OS Lactobacillus paracasei.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;

CC Lactobacillus.

OX NCBI\_TaxID=1597;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VTT E93490;

RA Nakayama J., Akkermans A.D.L., de Vos W.M.;

RT "Genomic survey of three component regulatory system putatively

involved in quorum sensing.";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF322594; AAG48147.1; -.

SQ SEQUENCE 306 AA; 35484 MW; 1E123E79905BE756 CRC64;

Query Match 40.7%; Score 43.5; DB 2; Length 306;  
Best Local Similarity 44.0%; Pred. No. 47;  
Matches 11; Conservative 3; Mismatches 6; Indels 5; Gaps 1;

QY 2 EAGNERISGLKY----TQIDQV 21  
| | | | | : | : | | | | |  
Db 246 ERGTNERFRELRYFPKGTQDQV 270

RESULT 11  
ID Q9WZY8 PRELIMINARY; PRT; 643 AA.  
AC Q9WZY8;



RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosalé C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003594; AAF51703.1; -  
 DR FlyBase; FBgn037069; CG7658.  
 DR InterPro; IPR000618; -  
 DR Pfam; PF00379; insect\_cuticle; 1.  
 DR PRINTS; PR00947; CUTICLE.  
 DR PROSITE; PS00233; CUTICLE; 1.  
 SQ SEQUENCE 119 AA; 12674 MW; DED4AB2C3C694B45 CRC64;

Query Match 39.3%; Score 42; DB 5; Length 119;  
 Best Local Similarity 60.0%; Pred. No. 27;  
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 QGAGNFERISGLKY 15  
 Db 55 QGAGNVNGISGSSY 69

RESULT 15  
 Q9SRH5 PRELIMINARY; PRT; 276 AA.  
 AC Q9SRH5;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
 DE PUTATIVE PORIN.  
 GN T22N4.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,  
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W., Fraser C.M.;  
 RA "Arabidopsis thaliana chromosome III BAC T22N4 genomic sequence."  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC010676; AAF03498.1; -

DR InterPro; IPR001925; -  
 DR Pfam; PF01459; Euk\_porin; 1.  
 DR PROSITE; PS00558; EUKARYOTIC\_PORIN; 1.  
 KW Porin.  
 SQ SEQUENCE 276 AA; 29425 MW; 7F33F53B56E9765C CRC64;  
 Query Match 39.3%; Score 42; DB 10; Length 276;  
 Best Local Similarity 35.0%; Pred. No. 72;  
 Matches 7; Conservative 7; Mismatches 6; Indels 0; Gaps 0;  
 Qy 2 EAGNFERISGLKYTQIDQV 21  
 Db 151 KSGNFTKINAGLSFTKEDLI 170

Search completed: July 5, 2001, 11:50:49  
 Job time: 472 sec



Thu Jul 5 13:59:10 2001

us-09-462-480-8.rspt

Page 7

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:45:42 ; Search time 130.35 Seconds  
(without alignments)  
19,534 Million cell updates/sec

Title: US-09-462-480-7  
 Perfect score: 206  
 Sequence: 1 0EAANK0K0ELDGI2TNIRO.....VOY5RADEE000A1SSOMGE 42

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 6062398 residues

Total number of hits satisfying chosen parameters: 412676

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

Database :

A\_Geneseq\_0601.\*

1:	/SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2:	/SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3:	/SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4:	/SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5:	/SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6:	/SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7:	/SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8:	/SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9:	/SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10:	/SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11:	/SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12:	/SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13:	/SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14:	/SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15:	/SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16:	/SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17:	/SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18:	/SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19:	/SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20:	/SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21:	/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22:	/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	198	96.1	42	20	AAV03707	M. tuberculosis LH	
2	198	96.1	80	18	AAW32454	Mycobacterium tube	
3	198	96.1	80	18	AAW32386	Mycobacterium tube	
4	198	96.1	80	19	AAW81707	M. tuberculosis im	
5	198	96.1	80	19	AAW64340	Mycobacterium tube	
6	198	96.1	80	20	AAW39137	M. tuberculosis an	
7	198	96.1	80	20	AAW38994	M. tuberculosis re	
8	198	96.1	95	18	AAW32444	Mycobacterium tube	
9	198	96.1	95	18	AAW32376	Mycobacterium tube	
10	198	96.1	95	19	AAW81747	M. tuberculosis im	
11	198	96.1	95	19	AAW64321	Mycobacterium tube	

12	198	96.1	95	20	AA932097	Mycobacterium tube
13	198	96.1	95	20	AA939118	M. tuberculosis an
14	198	96.1	95	20	AA938981	M. tuberculosis re
15	198	96.1	100	19	AAW81706	M. tuberculosis im
16	198	96.1	100	19	AAW64339	Mycobacterium tub
17	198	96.1	100	20	AA939136	M. tuberculosis an
18	198	96.1	100	20	AA938993	M. tuberculosis re
19	198	96.1	100	20	AA903705	M. tuberculosis LH
20	198	96.1	100	22	AAW35218	M. tuberculosis RV3
21	198	96.1	100	22	AAW19845	Mycobacterium tube
22	198	96.1	802	19	AAW81746	M. tuberculosis fu
23	198	96.1	802	19	AAW64379	Mycobacterium anti
24	198	96.1	802	20	AA932063	Mycobacterium tube
25	198	96.1	802	20	AA939224	M. tuberculosis fu
26	198	96.1	802	20	AA939176	M. tuberculosis fu
27	198	96.1	802	20	AA939081	M. tuberculosis fus
28	198	96.1	802	20	AA939033	M. tuberculosis fu
29	137	66.5	28	20	AA903712	M. tuberculosis LH
30	122	59.2	28	18	AAW32460	Mycobacterium tube
31	122	59.2	28	19	AAW81698	M. tuberculosis il
32	122	59.2	28	19	AA939128	M. tuberculosis an
33	111	53.9	27	18	AAW32459	Mycobacterium tube
34	111	53.9	27	19	AAW81697	M. tuberculosis im
35	111	53.9	27	20	AA939127	M. tuberculosis an
36	78	37.9	16	20	AA903713	M. tuberculosis LH
37	70	34.0	16	20	AA903711	M. tuberculosis LH
38	60	29.1	1325	18	AAW19540	M. tuberculosis an
39	60	29.1	1325	20	AAW94391	M. tuberculosis an
40	59.5	28.9	276	20	AAW00153	Enterococcus faeca
41	59.5	28.9	303	20	AAW00152	Enterococcus faeca
42	59	28.6	27	18	AAW32458	Mycobacterium tube
43	59	28.6	27	19	AAW81696	M. tuberculosis im
44	59	28.6	27	20	AAW39126	M. tuberculosis an
45	57	27.7	585	20	AAW97757	S-region transfer

## ALIGNMENTS

RESULT	1	
AA03707	AA03707	
ID	AA03707 standard; Protein; 42 AA.	
XX	XX	
XX	AA03707;	
XX	XX	
DT	07-JUN-1999 (first entry)	
XX	XX	
DE	M. tuberculosis LHP polypeptide antigenic fragment.	
XX	XX	
KW	ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;	
KW	immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;	
KW	immune response.	
XX	XX	
OS	Mycobacterium tuberculosis.	
XX	XX	
PN	WO9904005-A1.	
XX	XX	
XX	28-JAN-1999.	
PD	XX	
XX	16-JUL-1998; 98WO-IB01091.	
XX	XX	
PR	16-JUL-1997; 97US-0052631.	
XX	XX	
XX	(INSP ) INST PASTEUR.	
PA	(STAT-) STATENS SERUM INST.	
XX	XX	
PI	Andersen P, Berthet F, Gicquel B, Rasmussen PB;	
XX	XX	
DR	WPI; 1999-132249/11.	
XX	XX	
PT	New nucleic acid containing regulator and LHP gene of Mycobacterium	
PT	tuberculosis - useful in vaccines, for diagnosis, and for expression	
PT	of heterologous proteins	

xx Claim 21; Page 64; 88pp; English.  
 xx The present invention is directed to a polynucleotide carrying the  
 CC regulatory expression signals of the ESAT-6 protein as well as an open  
 CC reading frame coding for an antigenic protein LHP from Mycobacterium  
 CC tuberculosis. Host cells comprising the polynucleotide are used for the  
 CC recombinant expression of the protein. The recombinant polypeptide can  
 CC be used as immunogens and vaccines, to protect against bacteria of the  
 CC M. tuberculosis complex in humans or animals (the vaccines may include  
 CC other immunogenic proteins of the bacteria or their fragments,  
 CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by  
 CC detection of specific antibodies. The regulatory region present in the  
 CC polynucleotide may be used to express almost any heterologous protein in  
 CC mycobacteria, particularly as a fusion with polyhistidine. The two  
 CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to  
 CC provide a synergistic increase in ability to induce a protective immune  
 CC response. Sequences AAY03706-713 represent antigenic fragments of the  
 CC LHP polypeptide.  
 xx Sequence 42 AA;  
 SQ

Query Match 96.1%; Score 198; DB 20; Length 42;  
 Best Local Similarity 97.6%; Pred. No. 5.5e-20;  
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QEAANKQKQELDGISTNIRAGVQYSRADEEQQALSSQMGEF 42  
 Db 1 qeaankqkqeldeistniragvgvysradeeqqalssqmgf 42

RESULT 2  
 AAW32454  
 ID AAW32454 standard; Protein; 80 AA.  
 AC AAW32454;  
 XX 09-JAN-1998 (first entry)  
 DT Mycobacterium tuberculosis antigen Tb38-1F3.  
 DE Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M.tuberculosis.  
 XX Mycobacterium tuberculosis.  
 OS Mycobacterium tuberculosis.  
 XX WO9709428-A2.  
 PN 13-MAR-1997.  
 PD 30-AUG-1996; 96WO-US14674.  
 XX 12-JUL-1996; 96US-0680574.  
 PR 01-SEP-1995; 95US-0523436.  
 PR 22-SEP-1995; 95US-0533634.  
 PR 22-MAR-1996; 96US-0620874.  
 PR 05-JUN-1996; 96US-0659683.  
 XX (CORI-) CORIXA CORP.  
 PA Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;  
 XX WPI; 1997-192903/17.  
 DR N-PSDB; AAT91526.  
 XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
 PT useful in vaccines for prevention or treatment of tuberculosis, also  
 PT for diagnosis  
 XX Example 3; Page 149; 168pp; English.

CC A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M.tuberculosis  
 CC antigen, Tb38-1F3. The immunogenic protein, and fusion proteins  
 CC containing one or more of the proteins or one of the proteins plus  
 CC ESAT-6, are useful in vaccines, preferably when formulated with a  
 CC non-specific adjuvant, to induce an immune response against  
 CC M.tuberculosis (for treatment or prevention).  
 XX Sequence 80 AA;  
 SQ

Query Match 96.1%; Score 198; DB 18; Length 80;  
 Best Local Similarity 97.6%; Pred. No. 1.2e-19;  
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QEAANKQKQELDGISTNIRAGVQYSRADEEQQALSSQMGEF 42  
 Db 39 qeaankqkqeldeistniragvgvysradeeqqalssqmgf 80

RESULT 3  
 AAW32386  
 ID AAW32386 standard; Protein; 80 AA.  
 XX AAW32386;  
 AC AAW32386;  
 DT 13-JAN-1998 (first entry)  
 XX Mycobacterium tuberculosis antigen Tb38-1F3.  
 DE Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M.tuberculosis.  
 XX Mycobacterium tuberculosis.  
 OS Mycobacterium tuberculosis.  
 XX WO9709429-A2.  
 PN 13-MAR-1997.  
 PD 30-AUG-1996; 96WO-US14675.  
 XX 12-JUL-1996; 96US-0680573.  
 PR 01-SEP-1995; 95US-0523435.  
 PR 22-SEP-1995; 95US-0532136.  
 PR 22-MAR-1996; 96US-0620280.  
 PR 05-JUN-1996; 96US-0658800.  
 XX (CORI-) CORIXA CORP.  
 PA Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;  
 XX WPI; 1997-192904/17.  
 DR N-PSDB; AAT91460.  
 XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 PT - useful for diagnosis of M. tuberculosis infection  
 PT Example 3; Page 163; 190pp; English.

CC A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M.tuberculosis  
 CC antigen, Tb38-1F3. The immunogenic polypeptide can be used to diagnose  
 CC M.tuberculosis infection by forming complexes with specific  
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
 CC polypeptide can be used as diagnostic primers or probes and agents  
 CC that bind to the antigen, especially monoclonal antibodies or  
 CC equivalent polyclonal antibodies, are also used for diagnosis.  
 XX

SQ Sequence 80 AA;

Query Match 96.1%; Score 198; DB 19; Length 80;  
Best Local Similarity 97.6%; Pred. No. 1.2e-19;  
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGISTNIRQAGVQYSRADDEQQALSSQMGF 42  
|||||  
Db 39 qeaankqkqeldeistnirqagvysradeeqqalssqmgf 80

RESULT 4

AAW81707  
ID AAW81707 standard; Protein; 80 AA.

XX

AC AAW81707;

XX 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide Tb38-1F3.

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX W09816646-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18293.

XX 13-MAR-1997; 97US-0818112.

XX 11-OCT-1996; 96US-0730510.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.

XX N-PSDB; AAV64509.

Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
to develop products for the detection of M. tuberculosis infection  
and for diagnosis, treatment and prevention of tuberculosis

XX Example 3B; Page 139-140; 230pp; English.

This sequence represents an immunogenic portion of a soluble  
Mycobacterium tuberculosis (MT) antigen which can be used in a method  
for inducing protective immunity against tuberculosis (TB). This  
sequence can be formulated into vaccines and/or pharmaceutical  
compositions for immunising against M. tuberculosis infection or may  
be used for the diagnosis of tuberculosis.

XX Sequence 80 AA;

Query Match

Best Local Similarity 96.1%; Score 198; DB 19; Length 80;  
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGISTNIRQAGVQYSRADDEQQALSSQMGF 42  
|||||  
Db 39 qeaankqkqeldeistnirqagvysradeeqqalssqmgf 80

RESULT 5

AAW64340

ID AAW64340 standard; Protein; 80 AA.

XX

AC AAW64340;

XX 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen Tb38-1F3.

XX Tuberculosis; infection; diagnosis; antigen; Tb38-1F3.

XX Mycobacterium tuberculosis strain H37Rv.

XX W09816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

XX 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

XX N-PSDB; AAV44400.

New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
to develop products for the detection of M. tuberculosis infection  
and diagnosis of tuberculosis

XX Example 3; Page 146; 250pp; English.

This polypeptide comprises Mycobacterium tuberculosis antigen  
Tb38-1F3. A DNA molecule (see AAV44400) coding for Tb38-1F3 was  
isolated from a M. tuberculosis strain H37Rv genomic library. The  
invention relates to compositions and methods for diagnosing  
tuberculosis. It provides polypeptides (see AAW64291-W64379)  
comprising an antigenic portion of a soluble M. tuberculosis  
antigen, or an immunogenic portion of an M. tuberculosis antigen,  
as well as DNA sequences encoding such polypeptides, recombinant  
expression vectors and transformed or transfected host cells. Also  
claimed are methods and diagnostic kits for detecting M.  
tuberculosis infection in a patient using these polypeptides,  
antibodies or oligonucleotide probes and primers, for the diagnosis  
of tuberculosis.

XX Sequence 80 AA;

SQ

Query Match

Best Local Similarity 96.1%; Score 198; DB 19; Length 80;

Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGISTNIRQAGVQYSRADDEQQALSSQMGF 42  
|||||  
Db 39 qeaankqkqeldeistnirqagvysradeeqqalssqmgf 80

RESULT 6

AAV39137

ID AAV39137 standard; Protein; 80 AA.

XX

AC AAV39137;

XX 05-NOV-1999 (first entry)

XX M. tuberculosis antigen Tb38-1F3 amino acid sequence.

KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
KW immune response; skin test.

XX

OS Mycobacterium tuberculosis.

XX WO9942076-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03268.

XX 05-MAY-1998; 98US-0072967.

XX 18-FEB-1998; 98US-0025197.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527409/44.

DR N-PSDB; AAZ19310.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions

XX Example 3; Page 134-135; 299pp; English.

XX The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to  
CC AAY39225 are used in the exemplification of the present invention.

XX Sequence 80 AA;

Query Match 96.1%; Score 198; DB 20; Length 80;

Best Local Similarity 97.6%; Pred. No. 1.2e-19;

Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGTNIRAGVQYSRADDEEQQAALSSQMGF 42

Db 39 qeaankqkqeldeistniragvqysradeeqqalssqmgf 80

RESULT 7

AAV38994

ID AAY38994 standard; Protein; 80 AA.

XX AAY38994;

XX 05-NOV-1999 (first entry)

XX M. tuberculosis recombinant antigen protein Tb38-1F3.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;  
KW vaccine; immunity.

XX Mycobacterium tuberculosis.

XX WO9942118-A2.

XX 26-AUG-1999.

XX 17-FEB-1999; 99WO-US03265.

XX 05-MAY-1998; 98US-0072596.

XX 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

XX

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1999-527416/44.

DR N-PSDB; AAZ19098.

XX New polypeptide comprising antigenic portions of M. tuberculosis

XX Example 3; Page 180; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding  
CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
CC polypeptides are useful for detecting M. tuberculosis infection in a  
CC biological sample by detecting antibodies which bind with the  
CC polypeptides, and are useful as vaccines for immunizing against  
CC M. tuberculosis infection. The new detection methods are needed as  
CC current vaccination strategies do not provide 100% immunity.

XX Sequence 80 AA;

Query Match 96.1%; Score 198; DB 20; Length 80;

Best Local Similarity 97.6%; Pred. No. 1.2e-19;

Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGTNIRAGVQYSRADDEEQQAALSSQMGF 42

Db 39 qeaankqkqeldeistniragvqysradeeqqalssqmgf 80

RESULT 8

AAW32444

ID AAW32444 standard; Protein; 95 AA.

XX AAW32444;

XX 09-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen Tb38-1.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX WO9709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14674.

XX 12-JUL-1996; 96US-0680574.

XX 01-SEP-1995; 95US-0523436.

XX 22-SEP-1995; 95US-053634.

XX 22-MAR-1996; 96US-0620874.

XX 05-JUN-1996; 96US-0659683.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;

PI Twardzik DR, Vedvick TH;

XX WPI; 1997-192903/17.

XX N-PSDB; AAT91509.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
PT useful in vaccines for prevention or treatment of tuberculosis, also  
PT for diagnosis

XX Example 3; Page 124; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb38-1. The immunogenic protein, and fusion proteins  
CC containing one or more of the proteins or one of the proteins plus  
CC ESAT-6, are useful in vaccines, preferably when formulated with a  
CC non-specific adjuvant, to induce an immune response against  
CC M.tuberculosis (for treatment or prevention).  
XX  
XX  
SQ Sequence 95 AA;

Query Match 96.1%; Score 198; DB 18; Length 95;  
Best Local Similarity 97.6%; Pred. No. 1.4e-19;  
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGISTNIRQAGVOYSRADDEQQOQALSSQMGF 42  
DB 54 qeaankqkqeldeistnirqagvqysradeeqqalssqmgf 95

RESULT 9  
AAW32376  
ID AAW32376 standard; Protein; 95 AA.

XX AAW32376;  
XX  
XX 13-JAN-1998 (first entry)  
DT

DE Mycobacterium tuberculosis antigen Tb38-1.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX WO9709429-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US14675.

XX 12-JUL-1996; 96US-0680573.

XX 01-SEP-1995; 95US-0523435.

XX 22-SEP-1995; 95US-0532136.

XX 22-MAR-1996; 96US-0620280.

XX 05-JUN-1996; 96US-0658800.

XX (CORI-) CORIXA CORP.

XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
PI Twardzik DR, Vedvick TH;

XX WPI; 1997-192904/17.

XX N-PSDB; AAT91445.

XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
PT - useful for diagnosis of M. tuberculosis infection  
PT  
XX  
PS Example 3; Page 136; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
CC its variant differing only in conservative substitutions and/or  
CC modifications). The present sequence represents a M.tuberculosis  
CC antigen, Tb38-1. The immunogenic polypeptide can be used to diagnose  
CC M.tuberculosis infection by forming complexes with specific  
CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
CC polypeptide can be used as diagnostic primers or probes and agents  
CC that bind to the antigen, especially monoclonal antibodies or  
CC equivalent polyclonal antibodies, are also used for diagnosis.

XX Sequence 95 AA;

Query Match 96.1%; Score 198; DB 18; Length 95;  
Best Local Similarity 97.6%; Pred. No. 1.4e-19;  
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGISTNIRQAGVOYSRADDEQQOQALSSQMGF 42  
DB 54 qeaankqkqeldeistnirqagvqysradeeqqalssqmgf 95

RESULT 10  
AAW81747  
ID AAW81747 standard; Protein; 95 AA.

XX AAW81747;

XX 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide Tb38-1.

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX WO9816646-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18293.

XX 13-MAR-1997; 97US-0818112.

XX 11-OCT-1996; 96US-0730510.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.

XX N-PSDB; AAV64491.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis

XX Example 3b; Page 117; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble  
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
CC for inducing protective immunity against tuberculosis (TB). This sequence  
CC can be formulated into vaccines and/or pharmaceutical compositions for  
CC immunising against M. tuberculosis infection or may be used for the  
CC diagnosis of tuberculosis.

XX Sequence 95 AA;

Query Match 96.1%; Score 198; DB 19; Length 95;  
Best Local Similarity 97.6%; Pred. No. 1.4e-19;  
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGISTNIRQAGVOYSRADDEQQOQALSSQMGF 42  
DB 54 qeaankqkqeldeistnirqagvqysradeeqqalssqmgf 95

RESULT 11  
AAW64321  
ID AAW64321 standard; Peptide; 95 AA.

XX  
AC AAW64321;

```

XX DT 09-NOV-1998 (first entry)
XX DE Mycobacterium tuberculosis antigen Tb38-1 peptide.
XX KW Tuberculosis; infection; diagnosis; antigen; Tb38-1.
XX OS Mycobacterium tuberculosis strain H37Rv.
XX PN WO9816645-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18214.
XX PR 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1998-251292/22.
XX DR N-PSDB; AAV44384.
XX PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX PT to develop products for the detection of M. tuberculosis infection
XX PT and diagnosis of tuberculosis
XX PS Example 3; Page 123; 250pp; English.
XX CC This is an antigenic portion of Mycobacterium tuberculosis antigen
XX CC Tb38-1. A DNA sequence (see AAV44384) coding for antigen Tb38-1 was
XX CC isolated from a M. tuberculosis strain H37Rv expression library
XX CC using sera from patients having pulmonary or pleural tuberculosis.
XX CC The invention relates to compositions and methods for diagnosing
XX CC tuberculosis. It provides polypeptides (see AAV44384)
XX CC comprising an antigenic portion of a soluble M. tuberculosis
XX CC antigen, or an immunogenic portion of an M. tuberculosis antigen,
XX CC as well as DNA sequences encoding such polypeptides, recombinant
XX CC expression vectors and transformed or transfected host cells. Also
XX CC claimed are methods and diagnostic kits for detecting M.
XX CC tuberculosis infection in a patient using these polypeptides,
XX CC antibodies or oligonucleotide probes and primers, for the diagnosis
XX CC of tuberculosis.
XX SQ Sequence 95 AA;

Query Match 96.1%; Score 198; DB 19; Length 95;
Best Local Similarity 97.6%; Pred. No. 1.4e-19;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGISTNRQAGVGYSRADPEQQQALSSQMGEF 42
DB 54 qeaankqkqeldeistnrqagvgysradeeqqqalssqmgf 95

RESULT 12
AAV32097
ID AAV32097 standard; Protein; 95 AA.
XX AC AAV32097;
XX DT 17-JAN-2000 (first entry)
XX DE Mycobacterium tuberculosis antigen Tb38-1.
XX KW Tuberculosis; antigen; fusion protein; Tb38-1; diagnosis; therapy;
XX KW vaccine; immunogen.
XX OS Mycobacterium tuberculosis.

```

```

XX PN WO9951748-A2.
XX PD 14-OCT-1999.
XX PF 07-APR-1999; 99WO-US07717.
XX PR 07-APR-1998; 98US-0056556.
XX PR 30-DEC-1998; 98US-0223040.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Alderson M, Campos-Neto A;
XX PI WPI; 1999-601610/51.
XX DR New fusion proteins useful for diagnosis, prevention and treatment of
XX PT tuberculosis -
XX PS Claim 1; Fig 4D; 83pp; English.
XX CC This sequence represents the Mycobacterium tuberculosis antigen
XX CC Tb38-1. The invention provides fusion proteins (see AAV32059-71)
XX CC containing at least 2 M. tuberculosis antigens such as Tb38-1, e.g.
XX CC Tbf-2 (see AAV32063) and a TBH9-Tb38-1 fusion. The new fusion
XX CC proteins are useful as vaccines for preventing tuberculosis
XX CC (claimed), for diagnosis (via in vitro assays or intradermal skin
XX CC tests for detection of anti-M. tuberculosis antibodies), monitoring
XX CC of disease progression, and treatment of tuberculosis. They are
XX CC more effective immunogens than mixtures of the individual protein
XX CC components.
XX SQ Sequence 95 AA;

Query Match 96.1%; Score 198; DB 20; Length 95;
Best Local Similarity 97.6%; Pred. No. 1.4e-19;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEAANKQKQELDGISTNRQAGVGYSRADPEQQQALSSQMGEF 42
DB 54 qeaankqkqeldeistnrqagvgysradeeqqqalssqmgf 95

RESULT 13
AAV39118
ID AAV39118 standard; Protein; 95 AA.
XX AC AAV39118;
XX DT 05-NOV-1999 (first entry)
XX DE M. tuberculosis antigen Tb38-1 amino acid sequence.
XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;
XX KW immune response; skin test.
XX OS Mycobacterium tuberculosis.
XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US03268.
XX PR 05-MAY-1998; 98US-0072967.
XX PR 18-FEB-1998; 98US-0025197.
XX PA (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

```



XX WPI; 1999-527409/44.  
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 XX  
 PS Example 3; Page 113; 299pp; English.  
 XX The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to  
 CC AA219225 are used in the exemplification of the present invention.  
 XX  
 SQ Sequence 95 AA;  
 Query Match 96.1%; Score 198; DB 20; Length 95;  
 Best Local Similarity 97.6%; Pred. No. 1.4e-19;  
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QEAANKQKQELDGISTNIRQAGVQYSGRADEEQQALSSQMGF 42  
 |||||  
 DB 54 qeaankqkqeldestniragvgysradeeqqalssqmgf 95  
 RESULT 14  
 AAY38981  
 ID AAY38981 standard; Protein; 95 AA.  
 XX  
 AC AAY38981;  
 XX  
 DT 05-NOV-1999 (first entry)  
 XX  
 DE M. tuberculosis recombinant antigen protein Tb38-1.  
 XX Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN W09942118-A2.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 17-FEB-1999; 99WO-US03265.  
 XX  
 PR 05-MAY-1998; 98US-0072596.  
 PR 18-FEB-1998; 98US-0024753.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 WPI; 1999-527416/44.  
 DR N-PSDB; AA219082.  
 XX  
 PT New polypeptide comprising antigenic portions of M. tuberculosis  
 XX  
 PS Example 3; Page 159; 323pp; English.  
 XX This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against

CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.  
 XX  
 SQ Sequence 95 AA;  
 Query Match 96.1%; Score 198; DB 20; Length 95;  
 Best Local Similarity 97.6%; Pred. No. 1.4e-19;  
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QEAANKQKQELDGISTNIRQAGVQYSGRADEEQQALSSQMGF 42  
 |||||  
 DB 54 qeaankqkqeldestniragvgysradeeqqalssqmgf 95  
 RESULT 15  
 AA81706  
 ID AA81706 standard; Protein; 100 AA.  
 XX  
 AC AA81706;  
 XX  
 DT 27-JAN-1999 (first entry)  
 XX  
 DE M. tuberculosis immunogenic polypeptide Tb38-IN.  
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN W09816646-A2.  
 XX  
 PD 23-APR-1998.  
 XX  
 PF 07-OCT-1997; 97WO-US18293.  
 XX  
 PR 13-MAR-1997; 97US-0818112.  
 PR 11-OCT-1996; 96US-0730510.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 WPI; 1998-261042/23.  
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX  
 PS Example 3B; Page 138-139; 230pp; English.  
 XX This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.  
 XX  
 SQ Sequence 100 AA;  
 Query Match 96.1%; Score 198; DB 19; Length 100;  
 Best Local Similarity 97.6%; Pred. No. 1.5e-19;  
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 QEAANKQKQELDGISTNIRQAGVQYSGRADEEQQALSSQMGF 42  
 |||||  
 DB 59 qeaankqkqeldestniragvgysradeeqqalssqmgf 100  
 Search completed: July 5, 2001, 11:45:43

Job time: 272 sec

3





```

; NAME: Sherwood, Pamela
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: 06519/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-554-612C-50

```

```
Query Match      24.8%; Score 51; DB 1; Length 366;
Best Local Similarity 34.4%; Pred. NO. 12;
Matches 11; Conservative 7; Mismatches 14; Indels 0; Gaps 0;
```

RESULT 6  
US-08-416-756A-2  
; Sequence 2, Application US/08416756A  
; Patent No. 5750369  
; GENERAL INFORMATION:  
; APPLICANT: Lake, Staffan  
; APPLICANT: Stjernschantz, Johan  
; TITLE OF INVENTION: DNA Encoding A Prostaglandin F2' Receptor, A  
; TITLE OF INVENTION: Host Cell Transformed Therewith and An Expression Product  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
; STREET: 1100 New York Avenue, N. W.

CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20005-3934  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/416.756A  
 FILING DATE: 13-APR-1990

CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: WO PCT/SE93/00789  
FILING DATE: 01-OCT-1993  
PRIOR APPLICATION DATA:  
FILING NUMBER: SE 920892-7  
FILING DATE: 03-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Cimbala, Michele A.

```
/ / REGISTRATION NUMBER: 337851  
/ / REFERENCE/DOCKET NUMBER: 1370.00700000  
/ / TELECOMMUNICATION INFORMATION:  
/ /     TELEPHONE: 202-371-2600  
/ /     TELEFAX: 202-371-2540  
/ /     INFORMATION FOR SEQ ID NO: 2:  
/ /         SEQUENCE CHARACTERISTICS:  
/ /             LENGTH: 369 amino acids  
/ /             TYPE: amino acid  
/ /             TOPOLOGY: linear  
/ / MOLECULE TYPE: protein  
US-08-416-756A-2
```

US-08-416-756A-2

Query Match 24.8%; Score 51; DB 1; Length 369;  
Best Local Similarity 34.4%; Pred. No. 12;  
Matches 11; Conservative 7; Mismatches 14; Indels

QY 10 ELDGISTNIRQAGVQYSRADEEQQALSSQMG 41  
|| | : : | | | : | | : |  
Db 337 ELSSIKNSLKVAAISESPAEEKENQOASSEAG 368

RESULT 7  
US-08-416-756A-5  
; Sequence 5, Application US/08416756A  
; Patent No. 5750369  
; GENERAL INFORMATION:

APPLICANT: Lake, Staffan  
APPLICANT: Stjernschantz, Johan  
TITLE OF INVENTION: DNA Encoding A Prostaglandin F2' Receptor, A  
TITLE OF INVENTION: Host Cell Transformed Therewith and An Expression Product  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 New York Avenue, N. W.  
CITY: Washington

STATE: DC  
 COUNTRY: USA  
 ZIP: 20005-3934  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/416,756A  
 FILING DATE: 13-APR-1990  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: WO PCT/SE93/00789  
 FILING DATE: 01-OCT-1993

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: SE 9202892-7  
 ; FILING DATE: 02-OCT-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 :

```

1 NAME: Cimbala, Michele A.
2 REGISTRATION NUMBER: 33,851
3 REFERENCE/DOCKET NUMBER: 1370.00700000
4 TELECOMMUNICATION INFORMATION:
5 TELEPHONE: 202-371-2600
6 TELEFAX: 202-371-2540
7 INFORMATION FOR SEQ ID NO: 5:
8 SEQUENCE CHARACTERISTICS:
9 LENGTH: 390 amino acids
10 TYPE: amino acid
11 STRANDEDNESS: not relevant
12 TOPOLOGY: linear
13 MOLECULE TYPE: peptide
14 US-08-416-756A-5

```

Query Match 24.8%; Score 51; DB 1; Length 390;  
Best Local Similarity 34.4%; Pred. No. 13;  
Matches 11; Conservative 7; Mismatches 14; Indels.

QY 10 ELDGISTNIRQAGVQYSRADEEQQALSSQMG 41  
|| | : : | | | : | | : |  
Db 358 ELSSIKNSLKVAAISESPAEEKENQOASSEAG 389

RESULT 8  
US-08-562-535C-4  
; Sequence 4, Application US/08562535C  
; Patent No. 5916791  
; GENERAL INFORMATION:

Patent No. 5916791  
GENERAL INFORMATION:

APPLICANT: Joseph Hirschberg, Tamar Lotan and Mark Harker  
TITLE OF INVENTION: Polynucleotide molecule from Haematococcus pluvialis encoding  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein  
STREET: 2940 Birchtree space lane  
CITY: Silver Spring  
STATE: Maryland  
COUNTRY: United States of America  
ZIP: 20906  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
COMPUTER: Twinhead Slimnote-890TX  
OPERATING SYSTEM: MS DOS version 6.2,  
OPERATING SYSTEM: Windows version 3.11  
SOFTWARE: Word for Windows version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/562,535C  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Friedman, Mark M.  
REGISTRATION NUMBER: 33,883  
REFERENCE/DOCKET NUMBER: 325/5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 972-3-5625553  
TELEFAX: 972-3-5625554  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-562-535C-4

Query Match 23.8%; Score 49; DB 2; Length 329;  
Best Local Similarity 35.3%; Pred. No. 20;  
Matches 12; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 2 EAANKQKQELDGISTNRQAGVQYSRADDEQQQA 35  
|| ::::: || :| ||| |||  
Db 17 EALKEKEVAGSSDVLRTWATQYSLPSESDAA 50

RESULT 9  
US-08-742-605D-4  
Sequence 4, Application US/08742605D  
Patent No. 5965795  
GENERAL INFORMATION:  
APPLICANT: Joseph Hirschberg, Tamar Lotan and  
APPLICANT: Mark Harker  
TITLE OF INVENTION: Polynucleotide molecule from  
TITLE OF INVENTION: Haematococcus pluvialis encoding a  
TITLE OF INVENTION: polypeptide having a betha-C-4-oxygenase  
TITLE OF INVENTION: activity for biotechnological production of  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein  
STREET: 2940 Birchtree space lane  
CITY: Silver Spring  
STATE: Maryland  
COUNTRY: United States of America  
ZIP: 20906  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
COMPUTER: Chicom NB5500/386SX  
OPERATING SYSTEM: MS DOS version 6.2,  
OPERATING SYSTEM: Windows version 3.11

APPLICANT: Joseph Hirschberg, Tamar Lotan and  
APPLICANT: Mark Harker  
TITLE OF INVENTION: Polynucleotide molecule from  
TITLE OF INVENTION: Haematococcus pluvialis encoding a  
TITLE OF INVENTION: polypeptide having a betha-C-4-oxygenase  
TITLE OF INVENTION: activity for biotechnological production of  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mark M. Friedman c/o Robert Sheinbein  
STREET: 2940 Birchtree space lane  
CITY: Silver Spring  
STATE: Maryland  
COUNTRY: United States of America  
ZIP: 20906  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
COMPUTER: Chicom NB5500/386SX  
OPERATING SYSTEM: MS DOS version 6.2,  
OPERATING SYSTEM: Windows version 3.11

SOFTWARE: Word for Windows version 2.0,  
SOFTWARE: converted to ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/742,605D  
FILING DATE: Oct. 28, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/562,535  
FILING DATE: No. 5965795. 24, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Friedman, Mark M.  
REGISTRATION NUMBER: 33,883  
REFERENCE/DOCKET NUMBER: 325/12  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 972-3-5625553  
TELEFAX: 972-3-5625554  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 329 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-742-605D-4

Query Match 23.8%; Score 49; DB 2; Length 329;  
Best Local Similarity 35.3%; Pred. No. 20;  
Matches 12; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 2 EAANKQKQELDGISTNRQAGVQYSRADDEQQQA 35  
|| ::::: || :| ||| |||  
Db 17 EALKEKEVAGSSDVLRTWATQYSLPSESDAA 50

RESULT 10  
US-09-259-294-4  
Sequence 4, Application US/09259294  
Patent No. 6218599  
GENERAL INFORMATION:  
APPLICANT: Joseph Hirschberg, Tamar Lotan and  
APPLICANT: Mark Harker  
TITLE OF INVENTION: Polynucleotide molecule from  
TITLE OF INVENTION: Haematococcus pluvialis encoding a  
TITLE OF INVENTION: polypeptide having a betha-C-4-oxygenase  
TITLE OF INVENTION: activity for biotechnological production of  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mark M. Friedman c/o Anthony Castorina  
STREET: 2001 Jefferson Davis Highway, Suite 207  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
COMPUTER: Chicom NB5500/386SX  
OPERATING SYSTEM: MS DOS version 6.2,  
OPERATING SYSTEM: Windows version 3.11  
SOFTWARE: Word for Windows version 2.0,  
SOFTWARE: converted to ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/259,294  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: 08/742,605  
FILING DATE: Oct. 28, 1996  
APPLICATION NUMBER: 08/562,535  
FILING DATE: No. 6218599. 24, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Friedman, Mark M.  
REGISTRATION NUMBER: 33,883



;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/467,781  
;; FILING DATE: 06-JUN-1995  
;; CLASSIFICATION: 424  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: PITCHER ESQ, EDMUND R  
;; REGISTRATION NUMBER: 27,829  
;; REFERENCE/DOCKET NUMBER: MTP-013  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 248-7000  
;; TELEFAX: (617) 248-7100  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2101 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-467-781-4

Query Match 23.8%; Score 49; DB 1; Length 2101;  
Best Local Similarity 33.3%; Pred. No. 2.1e+02;  
Matches 10; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 6 KQKQELDGISTNIRQAGVQYSRADDEEQQA 35  
Db 459 EEKQOLSSLTLDQSSISNLSQAKELEQA 488

RESULT 14

US-08-195-487-4  
; Sequence 4, Application US/08195487  
; Patent No. 5783403  
; GENERAL INFORMATION:  
; APPLICANT: TOURKATLY, GARY  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 53 STATE STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/195,487  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION NUMBER: US/07/901,701  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/248-7000  
; TELEFAX: 617/248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-195-487-4

Query Match 23.8%; Score 49; DB 1; Length 2101;  
Best Local Similarity 33.3%; Pred. No. 2.1e+02;  
Matches 10; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 6 KQKQELDGISTNIRQAGVQYSRADDEEQQA 35  
Db 459 EEKQOLSSLTLDQSSISNLSQAKELEQA 488

RESULT 15

US-08-483-924-4  
; Sequence 4, Application US/08483924  
; Patent No. 5882876  
; GENERAL INFORMATION:  
; APPLICANT: TOURKATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,924  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-483-924-4

Query Match 23.8%; Score 49; DB 2; Length 2101;  
Best Local Similarity 33.3%; Pred. No. 2.1e+02;  
Matches 10; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

QY 6 KQKQELDGISTNIRQAGVQYSRADDEEQQA 35  
Db 459 EEKQOLSSLTLDQSSISNLSQAKELEQA 488

Search completed: July 5, 2001, 11:46:59  
Job time: 313 sec



7

THIS PAGE BLANK (USPTO)



1

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

```

Query Match          27.7%; Score 57; DB 1; Length 1961;
Best Local Similarity 36.1%; Pred. No. 82;
Matches 13; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

QY      4 ANKQKQLDGTNIRQGVQVYSRADEEQOQALSSQ 39
          : : : : : : : : : : : : : : : : : :
Db      1859 AEYKQDQADKASTRLKQLKQLEEAEEEAQGANASR 1894
          : : : : : : : : : : : : : : : : : :

RESULT 11
S21801
N: myosin heavy chain, neuronal [similarity] - rat
N: Alternate names: myosin II
N: Contains: myosin ATPase (EC 3.6.1.32)
C: Species: Rattus norvegicus (Norway rat)
C: Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
C: Accession: S21801; P00013; S18134
R: Sun, W.; Chantlier, P.D.
J. Mol. Biol. 224, 1185-1193, 1992
A: Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain
A: Reference number: S21801; MUID:92235856
A: Accession: S21801
A: Molecule type: mRNA
A: Residues: 1-1999 <SUN>
A: Cross-references: EMBL:X62659
R: Sun, W.; Chantlier, P.D.
Biochem. Biophys. Res. Commun. 175, 244-249, 1991
A: Title: A unique cellular myosin II exhibiting differential expression in the cerebral
A: Reference number: P00013; MUID:91151356
A: Accession: P00013
A: Molecule type: mRNA
A: Residues: 1914-1998, 'I' <SU2>
A: Experimental source: brain
C: Superfamily: myosin heavy chain; myosin motor domain homology
C: Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide
F: 84-763/Domain: myosin motor domain homology <MMOT>

```

```

Query Match          27.2%; Score 56; DB 1; Length 265;
Best Local Similarity 33.3%; Pred. No. 12;
Matches 16; Conservative 9; Mismatches 15; Indels 8; Gaps 2;

QY      3 AANKKQELDGIISTNIR---QAQVQYSRAD---EEQQOALSOMGF 42
        :|:|||||::|||::|||::|||::|||::|||::|||::
Db      136 SADKKIKRELDVAVDKVKELTQGEGTVYVRPRGIFSETLTALSEKYGY 183

RESULT   13
H64769
N;Alternate names: protein-export membrane protein secD; secretion protein secD
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: H64769; JQ0696; S12301
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: H64769
A;Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A;Residues: 1-615 <BLAT>
A;Cross-references: GB:AE000147; GB:U00096; NID:g1786603; PIDN:AAC73511.1; PID:g17866
A;Experimental source: strain K-12, substrain MG1655
R;Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.
EMBO J. 9, 3209-3216, 1990
A;Title: The sec locus of E.coli codes for two membrane proteins required for protei
A;Reference number: JQ0693; MUID:91006014
A;Accession: JQ0696
```

A:Molecule type: DNA  
A:Residues: 1-77,'S',79-154,'A',156-615 <GAR>  
A:Cross-references: GB:X56175; NID:942929; PIDN:CAA39634.1; PID:g581230  
R:Gardel, C.; Johnson, K.; Jacq, A.; Beckwith, J.  
EMBO J. 9, 4205-4206, 1990  
A:Reference number: S12298; MUID:91065354  
A:Contents: erratum  
A:Accession: S12301  
A:Molecule type: DNA  
A:Residues: 1-77,'S',79-154,'A',156-615 <GA2>  
A:Cross-references: EMBL:X56175; NID:942929; PID:g581230  
R:Pogliano, K.J.; Beckwith, J.  
J. Bacteriol. 176, 804-814, 1994  
A:Title: Genetic and molecular characterization of the Escherichia coli secD operon and  
A:Reference number: A36969; MUID:94131960  
A:Contents: annotation; membrane topology  
C:Genetics:  
A:Gene: secD  
A:Start codon: GTG  
C:Complex: heterohexamer; chains secY (PIR:QOEC5Y), secE (PIR:VXECSE), secG (PIR:S40402)  
ded trimeric complex of secY, secE and secG and the peripheral secA protein; the protein  
associated to form SecYEGDyajC, the hexameric integral membrane domain of the pre-protein  
C:Function:  
A:Description: the secD protein is a transmembrane component of the protein export compl  
C:Superfamily: protein export membrane protein secD  
C:Keywords: inner membrane; protein export; transmembrane protein  
F:10-30/Domain: transmembrane #status predicted <TM1>  
F:31-455/Domain: periplasmic #status predicted <P1>  
F:456-472/Domain: transmembrane #status predicted <TM2>  
F:477-497/Domain: transmembrane #status predicted <TM3>  
F:498-501/Domain: periplasmic #status predicted <PP2>  
F:502-518/Domain: transmembrane #status predicted <TM4>  
F:564-580/Domain: transmembrane #status predicted <TM5>  
F:581-585/Domain: periplasmic #status predicted <PP3>  
F:586-605/Domain: transmembrane #status predicted <TM6>

Query Match 27.2%; Score 56; DB 1; Length 615;  
Best Local Similarity 29.0%; Pred. No. 31;  
Matches 9; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 7 QKQELDGISTNIRQAGVQYSGRADEEQQAALS 37

Db 147 QEQNIDSLRDLREKIPYTTVRKENNYGLS 177

RESULT 14

G85536

hypothetical protein secD [imported] - Escherichia coli (strain O157:H7)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: G85536  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85536  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-615 <STO>  
A:Cross-references: GB:AE005174; NID:g12513258; PIDN:AAG54755.1; GSPDB:GN00145; UWGP:205  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: secD  
C:Superfamily: protein export membrane protein secD

Query Match 27.2%; Score 56; DB 2; Length 615;  
Best Local Similarity 29.0%; Pred. No. 31;  
Matches 9; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 7 QKQELDGISTNIRQAGVQYSGRADEEQQAALS 37

Db 147 QEQNIDSLRDLREKIPYTTVRKENNYGLS 177

RESULT 15

T10623

hypothetical protein F21C20.160 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Oct-1999  
C:Accession: T10623  
R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lem  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z16991  
A:Accession: T10623  
A:Molecule type: DNA  
A:Residues: 1-416 <BEV>  
A:Cross-references: EMBL:AL080254; GSPDB:GN00062; ATSP:F21C20.160  
A:Experimental source: cultivar Columbia; BAC clone F21C20  
C:Genetics:  
A:Gene: ATSP:F21C20.160  
A:Map position: 4  
A:Introns: 41/2; 81/3; 109/3; 146/2; 258/3; 326/3; 389/1

Query Match 26.9%; Score 55.5; DB 2; Length 416;  
Best Local Similarity 33.3%; Pred. No. 23;  
Matches 12; Conservative 12; Mismatches 11; Indels 1; Gaps 1;

QY 1 QEAANKQKQELD-GISTNIRQAGVQYSGRADEEQQA 35

Db 350 QESAGKVTTDIELATSSSDRQVGKSKREEEEEEA 385

Search completed: July 5, 2001, 11:48:32  
Job time: 371 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:51:42 ; Search time 41.8 seconds  
(without alignments)

34.419 Million cell updates/sec

Title: US-09-462-480-7

Perfect score: 206

Sequence: 1 QEAANKQKQELDGISTNIRQ.....VOYSRADDEQQQLSSQMGF 42

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	65	31.6	2022	1	ANT1_ONCVO
2	61	29.6	393	1	HEMX_ECOLI
3	60	29.1	1325	1	G160_MOUSE
4	57	27.7	1960	1	MYSN_HUMAN
5	56	27.2	265	1	YF02_BACST
6	56	27.2	615	1	SECD_ECOLI
7	56	27.2	615	1	SECD_SALCH
8	54.5	26.5	1033	1	Y328_MYCPN
9	54	26.2	1627	1	TP2B_CHICK
10	53.5	26.0	80	1	Y0C1_ANASP
11	53	25.7	468	1	Y0C1_CAEEL
12	53	25.7	576	1	Y0C1_CAEEL
13	53	25.7	591	1	DNAX_MYCCA
14	53	25.7	1552	1	TP2A_CHICK
15	52.5	25.5	382	1	IPAC_SHIDY
16	52.5	25.5	382	1	IPAC_SHIFL
17	52.5	25.5	494	1	MOT1_CRILLO
18	52	25.2	292	1	BPHC_PRESI
19	52	25.2	319	1	GDA5_WHEAT
20	52	25.2	992	1	EVC_HUMAN
21	51.5	25.0	78	1	Y0C1_SYNEL
22	51.5	25.0	80	1	Y0C1_MASLA
23	51.5	25.0	193	1	GVPC_ANAFL
24	51.5	25.0	433	1	MSRE_BOVIN
25	51.5	25.0	480	1	CFLA_HUMAN
26	51.5	25.0	636	1	FTHS_SPIOL
27	51.5	25.0	933	1	OD01_COXBU
28	51.5	25.0	1976	1	MYSO_HUMAN
29	51	24.8	248	1	CLCA_RAT
30	51	24.8	366	1	PF2R_MOUSE
31	51	24.8	366	1	PF2R_RAT
32	51	24.8	747	1	YMHA_CAEEL
33	51	24.8	822	1	MUS2_SYNY3

34 51 24.8 935 1 KINH\_SYNRA  
35 50 24.3 435 1 CLCA\_BOVIN  
36 50 24.3 435 1 AP50\_HUMAN  
37 50 24.3 462 1 VP33\_BPAPS  
38 50 24.3 771 1 DNK3\_SYNY3  
39 50 24.3 831 1 41\_HUMAN  
40 49.5 24.0 227 1 AG17\_ARATH  
41 49.5 24.0 455 1 NCBL\_MOUSE  
42 49.5 24.0 459 1 NCBL\_RAT  
43 49.5 24.0 554 1 MPIP\_YEAST  
44 49 23.8 235 1 CLCA\_MOUSE  
45 49 23.8 248 1 CLCA\_HUMAN

O43093 syncephalas  
P04973 bos taurus  
P20172 homo sapien  
Q97195 bacterioph  
P73098 synechocyst  
P11171 homo sapien  
Q38840 arabidopsis  
Q02819 mus musculu  
Q63083 rattus norv  
P23748 saccharomyc  
O08585 mus musculu  
P09496 homo sapien

#### ALIGNMENTS

RESULT 1  
ANT1\_ONCVO  
ID ANT1\_ONCVO STANDARD; PRT; 2022 AA.  
AC P21249;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE MAJOR ANTIGEN.  
GN OVT1.  
OS Onchocerca volvulus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
OC Onchocercidae; Onchocerca.  
OX NCBI\_TaxID=6282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95287898; PubMed=7770081;  
RA Tritereprapab S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,  
RA Neubert T.A., Scott A.L.;  
RT "Molecular cloning of a gene expressed during early embryonic  
development in Onchocerca volvulus."  
RL Mol. Biochem. Parasitol. 69:161-171(1995).  
RN [2]  
RP SEQUENCE OF 733-866 FROM N.A.  
RX MEDLINE=89127417; PubMed=2464764;  
RA Donelson J.E., Duke B.O.L., Moser D., Zeng W., Brundu N.E.,  
RA Lucius R., Renz A., Karam M., Flores G.Z.;  
RT "Construction of Onchocerca volvulus cDNA libraries and partial  
characterization of the cDNA for a major antigen."  
RL Mol. Biochem. Parasitol. 31:241-250(1988).  
CC -!- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; U12681; AAA80009.1; -;  
DR EMBL; J03995; AAA29412.1; -;  
DR PIR; A54513; A54513.  
DR HSSP; P02633; 3ICB.  
KW Antigen; Coiled coil.  
FT DOMAIN 74 120 COILED COIL (POTENTIAL).  
FT DOMAIN 151 251 COILED COIL (POTENTIAL).  
FT DOMAIN 327 384 COILED COIL (POTENTIAL).  
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).  
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;

Query Match 31.6%; Score 65; DB 1; Length 2022;  
Best Local Similarity 30.8%; Pred. No. 5;  
Matches 12; Conservative 15; Mismatches 15; Indels 0; Gaps 0;

OY 2 EAANKQKQELDGISTNIRQAGVQVSRADDEQQQLSSQMGF 40



01-OCT-2000 (Rel. 40, Last annotation update)  
 MYOSIN HEAVY CHAIN, NONMUSCLE TYPE A (CELLULAR MYOSIN HEAVY CHAIN,  
 TYPE A) (NMHC-A).  
 GN MYH9.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lloyd D.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-1337 FROM N.A.  
 RX MEDLINE=92003925; PubMed=1912569;  
 RA Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,  
 RA Arnout M.A., Clayton L.K., Tenen D.G.;  
 RT "Cellular myosin heavy chain in human leukocytes: isolation of 5'  
 RT cDNA clones, characterization of the protein, chromosomal  
 RT localization, and upregulation during myeloid differentiation.";  
 RN Blood 78:1826-1833(1991).  
 RN [3]  
 RP SEQUENCE OF 1-715 FROM N.A.  
 RX MEDLINE=91316803; PubMed=1860190;  
 RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,  
 RA Gdula D., Adelstein R.S., Weir L.;  
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located  
 RT on different chromosomes.";  
 RL Circ. Res. 69:530-539(1991).  
 RN [4]  
 RP SEQUENCE OF 714-1960 FROM N.A.  
 RX MEDLINE=90138958; PubMed=1967836;  
 RA Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;  
 RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity  
 RT through alternative polyadenylation.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).  
 CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
 CC CAPPING.  
 CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2  
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -!- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE  
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
 CC SUBFRAGMENT (S2).  
 CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY. STRONGEST  
 CC TO OTHER NONMUSCLE MYOSINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; 282215; CAB05105.1;  
 DR EMBL; M81105; AAA59888.1;  
 DR EMBL; M69180; AAA61765.1;  
 DR EMBL; M31013; AAA36349.1;  
 DR HSSP; P08799; LMND.  
 DR MIM; 160775;  
 DR InterPro; IPR000048;  
 DR InterPro; IPR001609;  
 DR InterPro; IPR002928;  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.

DR PROSITE; PS50096; IQ; 1.  
 KW Myosin; Coiled coil; Actin-binding; Alkylation; ATP-binding;  
 KW Multigene family; Calmodulin-binding.  
 FT DOMAIN 1 836 GLOBULAR HEAD (S1).  
 FT DOMAIN 837 1960  
 FT DOMAIN 837 1960 RODLIKE TAIL (S2 AND LMM DOMAINS).  
 FT NP\_BIND 174 181 COILED COIL (POTENTIAL).  
 FT DOMAIN 654 676 ATP.  
 FT DOMAIN 786 800 ACTIN-BINDING.  
 FT DOMAIN 812 829 CALMODULIN-BINDING (BY SIMILARITY).  
 FT MOD\_RES 694 694 CALMODULIN-BINDING (BY SIMILARITY).  
 FT MOD\_RES 704 704 ALKYLATION (SH-1) (POTENTIAL).  
 FT CONFLICT 53 55 ALKYLATION (SH-2) (POTENTIAL).  
 FT CONFLICT 53 55 EAI -> RGH (IN REF. 3).  
 FT CONFLICT 660 660 T -> S (IN REF. 3).  
 FT CONFLICT 869 869 T -> M (IN REF. 4).  
 FT CONFLICT 931 931 C -> Y (IN REF. 4).  
 FT CONFLICT 1240 1241 KG -> GR (IN REF. 4).  
 FT CONFLICT 1350 1350 E -> EE (IN REF. 2).  
 FT CONFLICT 1764 1764 T -> A (IN REF. 2).  
 FT CONFLICT 1771 1771 S -> G (IN REF. 2).  
 SQ SEQUENCE 1960 AA; 226531 MW; 588F84BB8C106E6F CRC64;  
 Query Match 27.7%; Score 57; DB 1; Length 1960;  
 Best Local Similarity 36.1%; Pred. No. 45;  
 Matches 13; Conservative 8; Mismatches 15; Indels 0; Gaps 0;  
 QY 4 ANKQKQELDGISTNIRQAGVQYSRADDEQQQALSSQ 39  
 DB 1858 AEQYKDOADKASTRLKRLKRLQLEEEAEAEQAANASR 1893  
 RESULT 5  
 YF02\_BACST STANDARD; PRT; 265 AA.  
 ID YF02\_BACST  
 AC Q04729;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE HYPOTHETICAL 30.6 KDA PROTEIN IN FUMA 3' REGION PRECURSOR (EC 3.5.1.-)  
 DE (ORF2).  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VAR. NON-DIASTATICUS / DSM 2234;  
 RX MEDLINE=9332761; PubMed=8473853;  
 RA Reaney S.K., Bungard S.J., Guest J.R.;  
 RT "Molecular and enzymological evidence for two classes of fumarase in  
 RT Bacillus stearothermophilus (var. non-diastaticus).";  
 RL J. Gen. Microbiol. 139:403-416(1993).  
 CC -!- SIMILARITY: TO OTHER POLYSACCHARIDE DEACETYLASES.  
 CC -!- SIMILARITY: TO B. SUBTILIS YLXY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L05611; AAA72318.1;  
 DR InterPro; IPR002509;  
 DR Pfam; PF01522; Polysac\_Deacet; 1.  
 KW Hypothetical protein; Hydrolase; Signal.  
 FT SIGNAL 1 23 POTENTIAL.  
 FT CHAIN 24 265 HYPOTHETICAL 30.6 KDA PROTEIN IN FUMA 3'  
 FT REGION (ORF2).  
 SQ SEQUENCE 265 AA; 30583 MW; 731D5203871730F3 CRC64;

-!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSDUCTION APPARATUS  
 WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH.  
 -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.  
 -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.

---

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

RESULT 7
SECID_SALCH STANDARD; PRT; 615 AA.
Q92FF8;
30-MAY-2000 (Rel. 39, Created)
30-MAY-2000 (Rel. 39, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
PROTEIN-EXPORT MEMBRANE PROTEIN SECID.
SECID.
Salmonella cholerae-suis (Salmonella enterica).
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
NCBI_TaxID=591;
[1]
SEQUENCE FROM N.A.
Kuhn F.C., Gies A.J., Smeltzer M., Crupper S.S., Sobieski R.J.;
"Identification of secD gene from Salmonella choleraesuis.";
Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: INVOLVED IN PROTEIN EXPORT.
-!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
(BY SIMILARITY).
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
-!- SIMILARITY: BELONGS TO THE SECID/SECF FAMILY. SECID FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF100611; AAC8384.1;
Protein transport; Translocation; Transmembrane; Inner membrane.
TRANSMEM 10 30 POTENTIAL.
TRANSMEM 452 472 POTENTIAL.
TRANSMEM 504 524 POTENTIAL.

```



```

Db      395 QAAAYMQQMQRMEIRHQQAQQAHHQQAQQHQAQHGMGY 436
| | : | | : | : || : |||| : |||| ||||
| | : | | : | : || : |||| : |||| ||||

RESULT 12
YNI5_YEAST
ID YNI5_YEAST STANDARD; PRT; 576 AA.
AC F53838;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 65.0 KDA PROTEIN IN MET2-SEC2 INTERGENIC REGION.
GN YNL275W OR N0626.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RA Messenguy F., Dubois E., Vierendeels F., Scherens B., Pierard A.,
RA Glansdorff M.;
RL Submitted May-1996) to the EMBL/GenBank/DDBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ANION EXCHANGER FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Z71551; CAA96183.1; -
CC SGD; S0005219; YNL275W.
DR InterPro: IPR003020; -.
DR Pfam: PF00955; HC03_cotransp; 1.
KW Hypothetical protein; Transmembrane; Anion exchange.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 236 256 POTENTIAL.
FT TRANSMEM 275 295 POTENTIAL.
FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 439 459 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
FT TRANSMEM 519 539 POTENTIAL.
SQ SEQUENCE 576 AA; 65027 MW; 4EA3FFC89F66307A CRC64;

Query Match 25.7%; Score 53; DB 1; Length 576;
Best Local Similarity 36.1%; Pred. No. 37;
Matches 13; Conservative 6; Mismatches 11; Indels 6; Gaps 1;

QY 2 EAANKQKQELDGISTNROAGVOYSRADEEQQAALS 37
: | : ||| : | : | : ||| : ||| : ||
DB 22 QALERNDLED-----RESSVESRSDEESHEKLS 51

RESULT 13
DNAA_MYCCA
ID DNAA_MYCCA STANDARD; PRT; 591 AA.
AC P45958;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNAAK PROTEIN (HEAT SHOCK 70 KDA PROTEIN) (HSP70).
GN DNAAK.
OS Mycoplasma capricolum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Entomoplasmataceae.
```

[illegible]





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:50:47 ; Search time 123.78 seconds  
(without alignments)  
44.893 Million cell updates/sec

Title: US-09-462-480-7  
Perfect score: 206  
Sequence: 1 QEAANKQKQELDGISTNIRQ.....VQYSRADEEQQALLSQMGF 42

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL16.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	198	96.1	100	2	069739		069739 mycobacteri
2	76	36.9	100	2	033084		033084 mycobacteri
3	60	29.1	585	4	09UH65		09UH65 homo sapien
4	60	29.1	585	4	09P028		09P028 homo sapien
5	60	29.1	603	4	075135		075135 homo sapien
6	60	29.1	1447	11	09QYT3		09QYT3 mus musculu
7	60	29.1	1487	11	09QYT2		09QYT2 mus musculu
8	58	28.2	109	9	038554		038554 lactococcus
9	58	28.2	620	9	038234		038234 bacterioph
10	58	28.2	706	9	038305		038305 lactococcus
11	57	27.7	521	2	09RAB2		09RAB2 bacillus ha
12	57	27.7	585	11	088443		088443 mus musculu
13	57	27.7	625	6	002717		002717 bos taurus
14	57	27.7	729	2	09X624		09X624 bacillus st
15	57	27.7	1959	11	063731		063731 rattus norv
16	56	27.2	615	2	09XBN2		09XBN2 enterobacte
17	56	27.2	615	2	09RNL9		09RNL9 citrobacter
18	56	27.2	784	4	09H1K0		09H1K0 homo sapien
19	56	27.2	2756	10	091J60		091J60 arabidopsis

20	55.5	26.9	416	10	Q9SVG6	Q9svg6 arabidopsis
21	55.5	26.9	478	10	Q9STW1	Q9stw1 arabidopsis
22	55.5	26.9	650	5	Q96580	Q96580 spisula sol
23	55.5	26.9	1976	6	Q27991	Q27991 bos taurus
24	55	26.7	224	2	Q9KQZ4	Q9kqz4 vibrio chol
25	55	26.7	389	2	Q9L6Q4	Q9l6q4 salmonella
26	55	26.7	743	2	Q9PPY9	Q9ppy9 ureaplasma
27	55	26.7	1961	11	Q62812	Q62812 rattus norv
28	54.5	26.5	243	10	Q9SXV9	Q9sxv9 lithospermu
29	54.5	26.5	249	4	015413	015413 homo sapien
30	54.5	26.5	579	4	Q9PIT3	Q9pit3 homo sapien
31	54	26.2	441	2	Q55718	Q55718 synechocyst
32	54	26.2	511	10	Q9SN02	Q9sn02 arabidopsis
33	54	26.2	1456	5	Q9W3M3	Q9w3m3 drosophila
34	54	26.2	1581	4	Q92614	Q92614 homo sapien
35	53.5	26.0	1330	4	043241	043241 homo sapien
36	53.5	26.0	1976	11	Q9ULT0	Q9ult0 rattus norv
37	53	25.7	293	2	Q9RBT1	Q9rbt1 pseudomonas
38	53	25.7	490	5	Q9NGS2	Q9ngs2 caenorhabdi
39	53	25.7	535	3	Q9UVG9	Q9uvvg9 mucor circi
40	53	25.7	874	14	012003	012003 chimpanzee
41	53	25.7	886	2	Q9KC72	Q9kc72 bacillus ha
42	53	25.7	1586	13	Q9DFB7	Q9dfb7 gallus gall
43	53	25.7	1871	5	Q9NCL3	Q9nc13 drosophila
44	53	25.7	2501	5	Q9NCW7	Q9ncw7 drosophila
45	52.5	25.5	253	1	Q9HQ59	Q9hqs9 halobacteri

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	100 AA.
069739	069739			
AC	069739			
DT	01-AUG-1998 (TREMBLrel. 07, Created)			
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)			
DE	HYPOTHETICAL 10.8 KDA PROTEIN.			
GN	LHP OR RV3874 OR MTV027.09.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,			
RA	Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RL	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RA	Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;			
RT	"Promoter analysis of the M. tuberculosis orf1C gene encoding the			
RL	early secreted antigenic target 6 kDa (ESAT-6)."			
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AL022120; CAAL7966.1; -			
DR	EMBL: AF004671; AAC83445.1; -			
DR	Tuberculist: Rv3874; -			
KW	Hypothetical protein.			
SQ	SEQUENCE 100 AA; 10794 MW; 285F4FC96F55D194 CRC64;			

Query Match 96.1%; Score 198; DB 2; Length 100;  
Best Local Similarity 97.6%; Pred. No. 2.7e-18;  
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QEANKQKQELDGISTNIRQAGVQYSRADEQQQALSSOMGF 42  
|||||  
DB 59 QEANKQKQELDGISTNIRQAGVQYSRADEQQQALSSOMGF 100

RESULT 2  
O33084 ID O33084 PRELIMINARY; PRT; 100 AA.  
AC O33084;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE SIMILARITY TO SWALL.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Eiglmeier K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93188700; PubMed=8446027;  
RA Eiglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;  
RT "Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae."  
RL Mol. Microbiol. 7:197-206(1993).  
DR EMBL; Y14967; CAA75210.1; -;  
SQ SEQUENCE 100 AA; 10964 MW; 460EE12F876BC383 CRC64;

Query Match 36.9%; Score 76; DB 2; Length 100;  
Best Local Similarity 36.6%; Pred. No. 0.014;  
Matches 15; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

QY 2 EAMQDQIRQLSIVDKLNRSGNKTDTDEANQLLSKNF 100  
|||  
DB 60 EAMQDQIRQLSIVDKLNRSGNKTDTDEANQLLSKNF 100

RESULT 3  
Q9UH65 ID Q9UH65 PRELIMINARY; PRT; 585 AA.  
AC Q9UH65;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE SWAP-70.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Masat L., Caldwell J., Jessberger R., Herndier B., Wabl M.,  
RA Ferrick D.;  
RT "Association of SWAP-70 with the B cell antigen receptor complex."  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO EF-HAND FAMILY.  
DR EMBL; AF210818; AAF24486.1; -;  
DR InterPro; IPR001849; -;  
DR InterPro; IPR002017; -;  
DR InterPro; IPR002048; -;  
DR Pfam; PF00036; ehand; 1.  
DR Pfam; PF00169; PH; 1.  
DR PROSITE; PSS0003; PH\_DOMAIN; 1.  
DR SMART; SM00233; PH; 1.  
KW Calcium-binding.

SQ SEQUENCE 585 AA; 68997 MW; B42B63CF033E612F CRC64;  
Query Match 29.1%; Score 60; DB 4; Length 585;  
Best Local Similarity 37.5%; Pred. No. 11;  
Matches 15; Conservative 11; Mismatches 10; Indels 4; Gaps 2;

QY 2 EAANKQK-QELDGISTNIRQAGVQYSRADEQQQALSSOM 40  
:||||:|  
DB 345 QAANESKQOELEAVRKKLEAA---SRAAEKKRLQTV 381

RESULT 4  
Q9P0Z8 ID Q9P0Z8 PRELIMINARY; PRT; 585 AA.  
AC Q9P0Z8;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE SWAP-70 HOMOLOG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Monz D.W., Comtesse N.E., Heckel D.;  
RT "Human SWAP-70 homolog."  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO EF-HAND FAMILY.  
DR EMBL; AF134894; AAF61403.1; -;  
DR InterPro; IPR001849; -;  
DR InterPro; IPR002017; -;  
DR InterPro; IPR002048; -;  
DR Pfam; PF00036; ehand; 1.  
DR Pfam; PF00169; PH; 1.  
DR PROSITE; PSS0003; PH\_DOMAIN; 1.  
DR SMART; SM00233; PH; 1.  
KW Calcium-binding.  
SQ SEQUENCE 585 AA; 68998 MW; AEEB7915C324A12F CRC64;

Query Match 29.1%; Score 60; DB 4; Length 585;  
Best Local Similarity 37.5%; Pred. No. 11;  
Matches 15; Conservative 11; Mismatches 10; Indels 4; Gaps 2;

QY 2 EAANKQK-QELDGISTNIRQAGVQYSRADEQQQALSSOM 40  
:||||:|  
DB 345 QAANESKQOELEAVRKKLEAA---SRAAEKKRLQTV 381

RESULT 5  
O75135 ID O75135 PRELIMINARY; PRT; 603 AA.  
AC O75135;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE KIAA0640 PROTEIN (FRAGMENT).  
GN KIAA0640.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=98403880; PubMed=9734811;  
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,  
RA Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. X.  
The complete sequences of 100 new cDNA clones from brain which can  
code for large proteins in vitro."

RL DNA Res. 5:169-176(1998).  
 CC -I- SIMILARITY: TO EF-HAND FAMILY.  
 DR EMBL; AB014540; BAA31615.1; -  
 DR InterPro; IPR001849; -  
 DR InterPro; IPR002017; -  
 DR InterPro; IPR002048; -  
 DR Pfam; PF00036; pfam.1.  
 DR Pfam; PF00169; pfam.1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR SMART; SM00233; PH; 1.  
 KW Calcium-binding.  
 FT NON\_TER 1  
 SQ SEQUENCE 603 AA; 71020 MW; 83C1677AAC5CCFC6 CRC64;

Query Match 29.1%; Score 60; DB 4; Length 603;  
 Best Local Similarity 37.5%; Pred. No. 11;  
 Matches 15; Conservative 11; Mismatches 10; Indels 4; Gaps 2;

QY 2 EAANKKQ-QLDGLSTNIRQAGVQYSPADEEQQAALSSQM 40  
 :|||: |||: : : : ||| |||: : : :  
 Db 363 QANESKQQLAVRKLEEA---SRAEEKKRLQTV 399

## RESULT 6

ID Q9QVT3 PRELIMINARY; PRT; 1447 AA.  
 AC Q9QVT3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE MALE ENHANCED ANTIGEN 2/GOLGI AUTOANTIGEN, GOLGIN SUBFAMILY A, 3.  
 GN MEA2/GOLGA3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL6;  
 RA Matsukuma S.;

Query Match 29.1%; Score 60; DB 11; Length 1447;  
 Best Local Similarity 40.0%; Pred. No. 29;  
 Matches 16; Conservative 11; Mismatches 11; Indels 2; Gaps 2;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB029537; BAA86889.1; JOINED.  
 DR EMBL; AB029521; BAA86889.1; JOINED.  
 DR EMBL; AB029522; BAA86889.1; JOINED.  
 DR EMBL; AB029523; BAA86889.1; JOINED.  
 DR EMBL; AB029524; BAA86889.1; JOINED.  
 DR EMBL; AB029525; BAA86889.1; JOINED.  
 DR EMBL; AB029526; BAA86889.1; JOINED.  
 DR EMBL; AB029527; BAA86889.1; JOINED.  
 DR EMBL; AB029528; BAA86889.1; JOINED.  
 DR EMBL; AB029529; BAA86889.1; JOINED.  
 DR EMBL; AB029530; BAA86889.1; JOINED.  
 DR EMBL; AB029531; BAA86889.1; JOINED.  
 DR EMBL; AB029532; BAA86889.1; JOINED.  
 DR EMBL; AB029533; BAA86889.1; JOINED.  
 DR EMBL; AB029534; BAA86889.1; JOINED.  
 DR EMBL; AB029535; BAA86889.1; JOINED.  
 DR EMBL; AB029536; BAA86889.1; JOINED.  
 SQ SEQUENCE 1447 AA; 162775 MW; 53087BE7460B3E6F CRC64;

Query Match 29.1%; Score 60; DB 11; Length 1447;  
 Best Local Similarity 40.0%; Pred. No. 29;  
 Matches 16; Conservative 11; Mismatches 11; Indels 2; Gaps 2;

QY 1 QEAANKKQQLDGLSTNIRQAGVQYSPADEEQQAALSSQM 40  
 :|||: |||: : : : ||| |||: : : :  
 Db 389 QERAELOAQ-LAALSTRL-QAQVEHSHSQKQDLSSEV 426

## RESULT 7

ID Q9QYT2 PRELIMINARY; PRT; 1487 AA.  
 AC Q9QYT2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE MEA2/GOLGA3 PROTEIN.  
 GN MEA2/GOLGA3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL6;  
 RA Matsukuma S.;

RT "Mea2/Golga3 Gene Product is Indispensable for Spermatogenesis in the Mouse.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB029537; BAA86890.1; JOINED.  
 DR EMBL; AB029521; BAA86890.1; JOINED.  
 DR EMBL; AB029522; BAA86890.1; JOINED.  
 DR EMBL; AB029523; BAA86890.1; JOINED.  
 DR EMBL; AB029524; BAA86890.1; JOINED.  
 DR EMBL; AB029525; BAA86890.1; JOINED.  
 DR EMBL; AB029526; BAA86890.1; JOINED.  
 DR EMBL; AB029527; BAA86890.1; JOINED.  
 DR EMBL; AB029528; BAA86890.1; JOINED.  
 DR EMBL; AB029529; BAA86890.1; JOINED.  
 DR EMBL; AB029530; BAA86890.1; JOINED.  
 DR EMBL; AB029531; BAA86890.1; JOINED.  
 DR EMBL; AB029532; BAA86890.1; JOINED.  
 DR EMBL; AB029533; BAA86890.1; JOINED.  
 DR EMBL; AB029534; BAA86890.1; JOINED.  
 DR EMBL; AB029535; BAA86890.1; JOINED.  
 DR EMBL; AB029536; BAA86890.1; JOINED.  
 SQ SEQUENCE 1487 AA; 167174 MW; 5BADF716C251FBFF CRC64;

Query Match 29.1%; Score 60; DB 11; Length 1487;  
 Best Local Similarity 40.0%; Pred. No. 30;  
 Matches 16; Conservative 11; Mismatches 11; Indels 2; Gaps 2;

QY 1 QEAANKKQQLDGLSTNIRQAGVQYSPADEEQQAALSSQM 40  
 :|||: |||: : : : ||| |||: : : :  
 Db 429 QERAELOAQ-LAALSTRL-QAQVEHSHSQKQDLSSEV 466

## RESULT 8

ID Q38554 PRELIMINARY; PRT; 109 AA.  
 AC Q38554;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE PHAGE ANTIGENIC DETERMINANT (FRAGMENT).  
 OS Lactococcus lactis phage phi 197.  
 OC Viruses.  
 OX NCBI\_TaxID=12427;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92384563; PubMed=1514794;  
 RA Schouler C., Bouet C., Ritzenthaler P., Drouet X., Mata M.;  
 RT "Characterization of Lactococcus lactis phage antigens.";  
 RL Appl. Environ. Microbiol. 58:2479-2484(1992).  
 DR EMBL; S44859; AAB49318.1; -  
 FT NON\_TER 109  
 SQ SEQUENCE 109 AA; 11900 MW; D8F90EB62539DF7B CRC64;

Query Match 28.2%; Score 58; DB 9; Length 109;  
 Best Local Similarity 37.5%; Pred. No. 3.2;



QY 4 ANKQKQELDGISTNIRQAGVQYSRADEEQQAALSSQ 39  
 \_ : \_ : \_ : \_ : \_ : \_ : \_ : \_ : \_ : \_ : \_ : \_ : \_ : \_ : \_ :  
 Db 518 AEYKQDQADKASTRLKQLKROLEAEAEAAQANASR 553

```

RESULT 15
Q63731 Q63731 PRELIMINARY; PRT; 1999 AA.
ID
AC Q63731;
AC Q63731;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE NEURONAL MYOSIN HEAVY CHAIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RP
RP TISSUE-BRAIN.
RC MEDLINE=92235856; PubMed=1569576;
RX
RA Sun W., Chantler P.D.;
RT "Cloning of the cDNA encoding a neuronal myosin heavy chain from
RT mammalian brain and its differential expression within the central
RT nervous system[see comments].";
RJ Mol. Biol. 224:1185-1193(1992).
RL

```

```

RESULT 15
Q63731
ID Q63731 PRELIMINARY; PRT; 1999 AA.
AC Q63731;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE NEURONAL MYOSIN HEAVY CHAIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OC NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RP
RP TISSUE=BRAIN;
RC
RC MEDLINE=92235856; PubMed=1569576;
RA Sun W., Chantler P.D.;
RX
RX "Cloning of the cDNA encoding a neuronal myosin heavy chain from
RT mammalian brain and its differential expression within the central
RT nervous system[see comments].";
RL J. Mol. Biol. 224:1185-1193(1992).
DR HSP; P10587; IBR1.
DR InterPro; IPR000048; -.
DR InterPro; IPR001609; -.
DR InterPro; IPR002017; -.
DR InterPro; IPR002928; -.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 2.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; -. 1.
DR SMART; SM00015; IQ; 1.
DR Myosin.
SQ
SQ SEQUENCE 1999 AA; 231456 MW; 090QC181D555727B41 CRC64;

```

Query Match	27.7%	Score 57;	DB 11;	Length 1999;
Best Local Similarity	36.1%	Pred. No. 1e+02;		
Matches 13; Conservative		8; Mismatches 15;	Indels 0;	Gaps 0;

QY 4 ANKQKQLDGI STNIRQAGVQYSRADEEQQA LSSQ 39  
| : | : | | : : | | : | : | : | :  
Db 1858 AEQYKDQADKA STRUKQLKRQLEEEEAQRANASR 1893

Search completed: July 5, 2001, 11:50:48  
Job time: 471 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:45:42 ; Search time 130.35 Seconds  
(without alignments)  
22.789 Million cell updates/sec

Title: US-09-462-480-6

Perfect score: 247

Sequence: 1 MAEMKTAATLQGEAGNFER.....DQVESTAGSLQGWRAAGT 49

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601:\*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	247	100.0	49	20	AAV03706	M. tuberculosis LH
2	247	100.0	100	20	AAV03705	M. tuberculosis LH
3	241	97.6	100	19	AAW81706	M. tuberculosis im
4	241	97.6	100	19	AAW64339	Mycobacterium tube
5	241	97.6	100	20	AAV39136	M. tuberculosis an
6	241	97.6	100	20	AAV38993	M. tuberculosis re
7	241	97.6	100	22	ABE35218	M tuberculosis RV3
8	241	97.6	100	22	ABE19845	Mycobacterium tube
9	236	95.5	802	19	AAW81746	M. tuberculosis fu
10	236	95.5	802	19	AAW64379	Mycobacterium anti
11	236	95.5	802	20	AAV32063	Mycobacterium tube

12	236	95.5	802	20	AAV39224	M. tuberculosis fu
13	236	95.5	802	20	AAV39176	M. tuberculosis fu
14	236	95.5	802	20	AAV39081	M. tuberculosis fu
15	236	95.5	802	20	AAV39033	M. tuberculosis fu
16	217	87.9	95	18	AAW32446	Mycobacterium tube
17	217	87.9	95	18	AAW32374	Mycobacterium tube
18	217	87.9	95	19	AAW81747	M. tuberculosis im
19	217	87.9	95	19	AAW64321	Mycobacterium tube
20	217	87.9	95	20	AAV32097	Mycobacterium tube
21	217	87.9	95	20	AAV39118	M. tuberculosis an
22	217	87.9	95	20	AAV38981	M. tuberculosis re
23	147	59.5	80	18	AAW32454	Mycobacterium tube
24	147	59.5	80	18	AAW32386	Mycobacterium tube
25	147	59.5	80	19	AAW81707	M. tuberculosis im
26	147	59.5	80	19	AAW64340	Mycobacterium tube
27	147	59.5	80	19	AAV39137	M. tuberculosis re
28	147	59.5	80	20	AAV38994	M. tuberculosis re
29	114	46.2	28	18	AAW32455	Mycobacterium tube
30	114	46.2	28	19	AAW81693	M. tuberculosis
31	114	46.2	28	20	AAV39123	M. tuberculosis LH
32	100	40.5	20	20	M. tuberculosis LH	Mycobacterium tube
33	85	34.4	16	18	AAW32456	M. tuberculosis im
34	85	34.4	16	19	AAW81694	M. tuberculosis im
35	85	34.4	16	20	AAV39124	Mycobacterium tube
36	84	34.0	27	18	AAW32457	Mycobacterium tube
37	84	34.0	27	19	AAW81695	M. tuberculosis im
38	84	34.0	27	20	AAV39125	M. tuberculosis im
39	78	31.6	16	20	AAV03709	M. tuberculosis LH
40	71	28.7	16	20	AAV03710	M. tuberculosis LH
41	63	25.5	103	22	AAV35234	M tuberculosis RV3
42	62	25.1	343	16	AAW82455	Hepatitis GB virus
43	62	25.1	343	21	AAW09478	Hepatitis GB virus
44	62	25.1	2972	21	AAW09265	Hepatitis GB virus
45	62	25.1	3163	16	AAW94347	Hepatitis GB virus

#### ALIGNMENTS

RESULT 1  
AAV03706  
ID AAV03706 standard; Protein; 49 AA.

XX AC AAV03706;

XX DT 07-JUN-1999 (first entry)

XX DE M. tuberculosis LHP polypeptide antigenic fragment.

XX ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
KW immune response.

XX OS Mycobacterium tuberculosis.

XX PN WO9904005-A1.

XX PD 28-JAN-1999..

XX PF 16-JUL-1998; 98WO-IB01091.

XX PR 16-JUL-1997; 97US-0052631.

XX PA (INSP ) INST PASTEUR.

XX PA (STAT-) STATENS SERUM INST.

XX PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX DR WPI; 1999-132249/11.

XX PT New nucleic acid containing regulator and LHP gene of Mycobacterium  
PT tuberculosis - useful in vaccines, for diagnosis, and for expression  
PT of heterologous proteins

XX PS Claim 21; Page 64; 88pp; English.

CC The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can

CC be used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. Sequences AAY03706-713 represent antigenic fragments of the

CC LHP polypeptide.

XX SQ Sequence 49 AA;

Query Match 100.0%; Score 247; DB 20; Length 49;

Best Local Similarity 100.0%; Pred. No. 7.2e-26;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGQAGNFERISGLKTKTQIDQVESTAGSLQGWGRGAAGT 49

DB 1 maemktdaatlgqagnferisgdlktqldqvvestagslqgqwrigaagt 49

RESULT 2

AAY03705

ID AAY03705 standard; Protein; 100 AA.

XX AC AAY03705;

XX DT 07-JUN-1999 (first entry)

XX DE M. tuberculosis LHP polypeptide.

KW ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;

KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;

KW immune response.

XX OS Mycobacterium tuberculosis.

XX PN WO9904005-A1.

XX PD 28-JAN-1999.

XX PF 16-JUL-1998; 98WO-IB01091.

XX PR 16-JUL-1997; 97US-0052631.

XX PA (INSP ) INST PASTEUR.

XX PA (STAT-) STATENS SERUM INST.

XX PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;

XX DR WPI; 1999-132249/11.

DR N-PSDB; AAX29168, AAX29171.

XX New nucleic acid containing regulator and LHP gene of Mycobacterium

PT tuberculosis - useful in vaccines, for diagnosis, and for expression

PT of heterologous proteins

XX Claim 17; Page 64; 88pp; English.

XX PS The present invention is directed to a polynucleotide carrying the

CC regulatory expression signals of the ESAT-6 protein as well as an open

CC reading frame coding for an antigenic protein LHP from Mycobacterium

CC tuberculosis. Host cells comprising the polynucleotide are used for the

CC recombinant expression of the protein. The recombinant polypeptide can be

CC used as immunogens and vaccines, to protect against bacteria of the

CC M. tuberculosis complex in humans or animals (the vaccines may include

CC other immunogenic proteins of the bacteria or their fragments,

CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by

CC detection of specific antibodies. The regulatory region present in the

CC polynucleotide may be used to express almost any heterologous protein in

CC mycobacteria, particularly as a fusion with polyhistidine. The two

CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to

CC provide a synergistic increase in ability to induce a protective immune

CC response. The present sequence represents the LHP polypeptide.

XX SQ Sequence 100 AA;

Query Match 100.0%; Score 247; DB 20; Length 100;

Best Local Similarity 100.0%; Pred. No. 1.7e-25;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGQAGNFERISGLKTKTQIDQVESTAGSLQGWGRGAAGT 49

DB 1 maemktdaatlgqagnferisgdlktqldqvvestagslqgqwrigaagt 49

RESULT 3

AAW81706

ID AAW81706 standard; Protein; 100 AA.

XX AC AAW81706;

XX DT 27-JAN-1999 (first entry)

XX DE M. tuberculosis immunogenic polypeptide TB38-IN.

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;

KW vaccine; pharmaceutical; infection; diagnosis.

XX OS Mycobacterium tuberculosis.

XX PN WO9816646-A2.

XX PD 23-APR-1998.

XX PF 07-OCT-1997; 97WO-US18293.

XX PR 13-MAR-1997; 97US-0818112.

XX PR 11-OCT-1996; 96US-0730510.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX DR WPI; 1998-261042/23.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used

PT to develop products for the detection of M. tuberculosis infection

PT and for diagnosis, treatment and prevention of tuberculosis

XX Example 3B; Page 138-139; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble

CC Mycobacterium tuberculosis (MT) antigen which can be used in a method

CC for inducing protective immunity against tuberculosis (TB). This

CC sequence can be formulated into vaccines and/or pharmaceutical

CC compositions for immunising against M. tuberculosis infection or may

CC be used for the diagnosis of tuberculosis.

XX SQ Sequence 100 AA;

Query Match 97.6%; Score 241; DB 19; Length 100;



Best Local Similarity 98.0%; Pred. No. 1.1e-24;  
Matches 48; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGQWRGAAGT 49  
Db 1 maemktdaatlaaqnferisgdlktqldqvvestagslqqwrgaagt 49

RESULT 4  
AAW64339 standard; Protein; 100 AA.  
XX AC AAW64339;  
XX DT 09-NOV-1998 (first entry)  
XX DE Mycobacterium tuberculosis antigen Tb38-IN.  
XX KW Tuberculosis; infection; diagnosis; antigen; Tb37-FL.  
XX OS Mycobacterium tuberculosis strain H37Rv.  
XX PN W09816645-A2.  
XX PD 23-APR-1998.  
XX PF 07-OCT-1997; 97WO-US18214.  
XX PR 13-MAR-1997; 97US-0818111.  
XX PR 11-OCT-1996; 96US-0729622.  
XX PA (CORI-) CORIXA CORP.  
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
XX WPI; 1998-251292/22.  
XX PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and diagnosis of tuberculosis  
XX Example 3; Page 145; 250pp; English.  
XX This polypeptide comprises a partial sequence of Mycobacterium  
CC tuberculosis antigen Tb38-IN. It is encoded by genomic DNA isolated  
CC from a M. tuberculosis strain H37Rv genomic library using a probe  
CC derived from clone Tb38-1 (see AAW44384). The invention relates to  
CC compositions and methods for diagnosing tuberculosis. It provides  
CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of  
CC a soluble M. tuberculosis antigen, or an immunogenic portion of an  
CC M. tuberculosis antigen, as well as DNA sequences encoding such  
CC polypeptides, recombinant expression vectors and transformed or  
CC transfected host cells. Also claimed are methods and diagnostic  
CC kits for detecting M. tuberculosis infection in a patient using  
CC these polypeptides, antibodies or oligonucleotide probes and  
CC primers, for the diagnosis of tuberculosis.  
XX Sequence 100 AA;

Query Match 97.6%; Score 241; DB 19; Length 100;  
Best Local Similarity 98.0%; Pred. No. 1.1e-24;  
Matches 48; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGQWRGAAGT 49  
Db 1 maemktdaatlaaqnferisgdlktqldqvvestagslqqwrgaagt 49

RESULT 5  
AAW39136 standard; Protein; 100 AA.  
ID AAW39136

XX AAY39136;  
XX AC 05-NOV-1999 (first entry)  
XX DT M. tuberculosis antigen Tb38-IN amino acid sequence.  
XX DE Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
XX KW immune response; skin test.  
XX OS Mycobacterium tuberculosis.  
XX PN W09942076-A2.  
XX PD 26-AUG-1999.  
XX PF 17-FEB-1999; 99WO-US03268.  
XX PR 05-MAY-1998; 98US-0072967.  
XX PR 18-FEB-1998; 98US-0025197.  
XX PA (CORI-) CORIXA CORP.  
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
XX WPI; 1999-527409/44.  
XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions  
XX Example 3; Page 133-134; 299pp; English.  
XX The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to  
CC AAY39225 are used in the exemplification of the present invention.  
XX Sequence 100 AA;

Query Match 97.6%; Score 241; DB 20; Length 100;  
Best Local Similarity 98.0%; Pred. No. 1.1e-24;  
Matches 48; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGQWRGAAGT 49  
Db 1 maemktdaatlaaqnferisgdlktqldqvvestagslqqwrgaagt 49

RESULT 6  
AAW38993 standard; Protein; 100 AA.  
ID AAW38993  
XX AC AAW38993;  
XX DT 05-NOV-1999 (first entry)  
XX DE M. tuberculosis recombinant antigen protein Tb38-IN.  
XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
XX KW vaccine; immunity.  
XX OS Mycobacterium tuberculosis.  
XX



Query Match 97.6%; Score 241; DB 22; Length 100;  
 Best Local Similarity 98.0%; Pred. No. 1.1e-24;  
 Matches 48; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLQGEAGNFERISGDLKTDQVSTAGSLQGQWRGAAGT 49  
 |||||  
 Db 1 maemktdaatlaqaegnferisgdlktqtdqvstagslqgqwrngaagt 49

## RESULT 9

AAW81746  
 ID AAW81746 standard; Protein; 802 AA.

XX AC AAW81746;

DT 27-JAN-1999 (first entry)

DE M. tuberculosis fusion protein Tbf-2.

KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.

OS Synthetic.

OS Mycobacterium tuberculosis.

PN WO9816646-A2.

XX PD 23-APR-1998.

XX PF 07-OCT-1997; 97WO-US18293.

PR 13-MAR-1997; 97US-0818112.

PR 11-OCT-1996; 96US-0730510.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;

XX DR WPI: 1998-261042/23.

XX DR N-PSDB; AAW64379.

XX This sequence represents the fusion protein Tbf-2 which is composed of  
 CC immunogenic polypeptides from Mycobacterium tuberculosis (MT). This  
 CC protein is used in a method for inducing protective immunity against  
 CC tuberculosis (TB). This sequence can be formulated into vaccines  
 CC and/or pharmaceutical compositions for immunising against  
 CC M. tuberculosis infection or may be used for the diagnosis of TB.

XX SQ Sequence 802 AA;

XX Query Match 95.5%; Score 236; DB 19; Length 802;  
 XX Best Local Similarity 97.9%; Pred. No. 6.3e-23;  
 XX Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTDAAATLQGEAGNFERISGDLKTDQVSTAGSLQGQWRGAAGT 49  
 |||||  
 Db 425 aemktdaatlaqaegnferisgdlktqtdqvstagslqgqwrngaagt 472

## RESULT 10

AAW64379  
 ID AAW64379 standard; Protein; 802 AA.

XX

AC AAW64379;

XX 09-NOV-1998 (first entry)

XX Mycobacterium antigen Tbf2 protein fusion.

XX Tuberculosis; infection; diagnosis; 38 kDa antigen; Tbra3; DPEP;  
 KW Tbf38-1; Tbf-2.

XX OS Mycobacterium tuberculosis.

OS Synthetic.

PN WO9816645-A2.

XX PD 23-APR-1998.

XX PF 07-OCT-1997; 97WO-US18214.

PR 13-MAR-1997; 97US-0818111.

PR 11-OCT-1996; 96US-0729622.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedwick TS;

XX DR WPI: 1998-251292/22.

XX DR N-PSDB; AAV55801.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 XX to develop products for the detection of M. tuberculosis infection  
 XX and diagnosis of tuberculosis

XX Example 7; Page 223-226; 250pp; English.

XX This polypeptide comprises a fusion protein, designated Tbf-2,  
 CC composed of Mycobacterium tuberculosis antigens Tbra3 (see AAW64295),  
 CC 38 kDa antigen (see AAW64364), Tbf38-1 (see AAW64321) and DPEP (see  
 CC AAW64322). It was produced by PCR amplification (see AAV44450-57) of  
 CC the appropriate antigen DNA sequences, cloning into an expression  
 CC vector, and expression in E. coli. Tbf-2 can be used for  
 CC serodiagnosis of tuberculosis. The invention relates to  
 CC compositions and methods for diagnosing tuberculosis. It provides  
 CC polypeptides (see AAW64291-W64379) comprising antigenic or  
 CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,  
 CC DNA sequences encoding such polypeptides, recombinant expression  
 CC vectors and host cells. Also claimed are methods and diagnostic  
 CC kits for detecting M. tuberculosis infection in a patient.

XX SQ Sequence 802 AA;

XX Query Match 95.5%; Score 236; DB 19; Length 802;  
 XX Best Local Similarity 97.9%; Pred. No. 6.3e-23;  
 XX Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTDAAATLQGEAGNFERISGDLKTDQVSTAGSLQGQWRGAAGT 49  
 |||||  
 Db 425 aemktdaatlaqaegnferisgdlktqtdqvstagslqgqwrngaagt 472

## RESULT 11

AAV32063  
 ID AAV32063 standard; Protein; 802 AA.

XX AC AAV32063;

XX DT 17-JAN-2000 (first entry)

XX Mycobacterium tuberculosis antigen fusion protein Tbf-2.

XX Tuberculosis; antigen; fusion protein; Tbf-2; Tbra3; 38kD; Tbf38-1;  
 KW DPEP; diagnosis; therapy; vaccine; immunogen.









```

3
RESULT
PCT-US96-06511-4
; Sequence 4, Application PC/TUS9606511
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel S.
; APPLICANT: Walker, John C.
; TITLE OF INVENTION: Glycine and Phaseolus
; TITLE OF INVENTION: alpha-D-Galactosidases
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 Northwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: US
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

4  
RESULT US-08-488-961-7  
.: Sequence 7, Application US/08488961  
.: Patent No. 5606042  
.: GENERAL INFORMATION:  
.: APPLICANT: Smith, Daniel S.  
.: APPLICANT: Walker, John C.  
.: TITLE OF INVENTION: Glycine and Phaseolus  
.: TITLE OF INVENTION: alpha-D-Galactosidases  
.: NUMBER OF SEQUENCES: 11  
.: CORRESPONDENCE ADDRESS:  
.: ADDRESSEE: Reising, Ethington, Barnard & Perry  
.: STREET: P.O. Box 4390  
.: CITY: Troy  
.: STATE: Michigan  
.: COUNTRY: US  
.: ZIP: 48099-4390  
.: COMPUTER READABLE FORM:  
.: MEDIUM TYPE: Floppy disk  
.: COMPUTER: IBM PC compatible  
.: OPERATING SYSTEM: PC-DOS/MS-DOS  
.: SOFTWARE: PatentIn Release #1.0, Version #1.30  
.: CURRENT APPLICATION NUMBER: US/08/488.961  
.: FILING DATE:  
.: CLASSIFICATION: 435  
.: ATTORNEY/AGENT INFORMATION:  
.: NAME: Kohn, Kenneth I.  
.: REGISTRATION NUMBER: 30,955  
.: REFERENCE/DOCKET NUMBER: P-320 (UMO)  
.: TELECOMMUNICATION INFORMATION:  
.: TELEPHONE: (810) 689-3500  
.: TELEFAX: (810) 689-4071  
.: INFORMATION FOR SEQ ID NO: 7:  
.: SEQUENCE CHARACTERISTICS:  
.: LENGTH: 363 amino acids  
.: TYPE: amino acid  
.: STRANDEDNESS: single  
.: TOPOLOGY: linear  
US-08-488-961-7

Query Match 22.9%; Score 56.5; DB 1; Length 363;  
Best Local Similarity 33.3%; Pred. No. 4.7;









```
; MOLECULE TYPE: protein
US-08-646-715-9

Query Match      21.7%; Score 53.5; DB 1; Length 278;
Best Local Similarity 26.9%; Pred. No. 8.7;
Matches 14; Conservative 9; Mismatches 22; Indels 7; Gaps 1;

Qy 5 KTDAAATLQGEAGNFERI-----SGDLKTQIDQVESTAGSLQGWGAAGT 49
Db 196 KTVSGSGGGGGGQEVKSESTGAGDLKMEVDSDAAVGSIAGSGSAGS 247

RESULT 13
US-08-452-592B-6
; Sequence 6, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; TITLE OF INVENTION: Self-Assembling Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: WordPerfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/068,948
; FILING DATE: 25 May 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Donahue, Richard J.
; NAME: Lamming, John H.
; NAME: Ranucci, Vincent J.
; REGISTRATION NUMBER: 22,062 (Donahue)
; REGISTRATION NUMBER: 34,857 (Lamming)
; REGISTRATION NUMBER: 29,579 (Ranucci)
; REFERENCE/DOCKET NUMBER: NA-1096D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-233-4510
; TELEFAX: 508-233-5167
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-452-592B-7

Query Match      21.1%; Score 52; DB 1; Length 84;
Best Local Similarity 30.8%; Pred. No. 3;
Matches 12; Conservative 10; Mismatches 17; Indels 0; Gaps 0;

Qy 2 AEMKTAATLQGEAGNFERISGDLKTQIDQVESTAGSLQ 40
Db 23 AELQEVSRLENEIEDLKAKIGDLKNKVAQLKRVSLK 61

RESULT 15
US-08-665-647-3
; Sequence 3, Application US/08665647
; Patent No. 5935803
; GENERAL INFORMATION:
; APPLICANT: Dasquez, Nicki J.
; APPLICANT: Ron, Dorit
; APPLICANT: Voronova, Anna F.
; APPLICANT: Napolitano, Eugene W.
; TITLE OF INVENTION: METHODS TO IDENTIFY IMMUNOMODULATORS
; TITLE OF INVENTION: USING COGNATE INTERACTION OF PKC-THETA
; NUMBER OF SEQUENCES: 89
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW - Ste. 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,647
; FILING DATE: 18-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 22550-20025.25
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 822-0168
; TELEX: 90-4030 MRSNFOERSH
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-665-647-3

```

```

Query Match      21.1%; Score 52; DB 2; Length 334;
Best Local Similarity 31.7%; Pred. No. 18;
Matches 13; Conservative 8; Mismatches 18; Indels 2; Gaps 1;
QY 7 DAATLGOEAGNFRISGDLKTQIDQVESTAGSLQGGWRGAA 47
Db 248 NAAEAGQEQAVIGLQELK--VEPLELAGVKKEEVWEGTA 286

```

Search completed: July 5, 2001, 11:46:58  
Job time: 312 sec

THIS PAGE BLANK (uspto)



**QY**     **1** MAEMKTDAAITLGQAGNPERISGLDKTQIDOVESTAGSLQCQWRGAAGT 49  
||||| :||| | :||| : ||| : ||| :

**Dd**       **1** MAEMITEAAILTOQAADFQTIASLSOEERNEFVDSIGOSFONTWEGQAAS 49

**RESULT**      **3**

B07600

hypothetical protein Rv3903c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: B07600

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; PMID:98295987

A:Accession: B07600

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-103 <COL>

A:Cross-references: GB:Z94121; GB:AL123456; NID:g3261736; PIDN:CAB08096.1; PTD:e312272;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: Rv3905c

```

Query Match      25.5%; Score 63; DB 2; Length 103;
Best Local Similarity 23.9%; Pred. NO. 1;
Matches 11; Conservative 15; Mismatches 20; Indels 0; Gaps 0;

QY 4 MKTDAATLGGGAGNFRISGLDKTQIDQVESTAGSLQGQWRGAAGT 49
      : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : :
Db 7 LRVEPVMQGFASLSDGAAHHLAVQLAEIDAQVGMILGGWRGASGS 52
      : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 4

D83205

probable oxidoreductase PA3534 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*

C:date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: D83205

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen

A:reference number: AB2950; MUID:20437337

A:Accession: D83205

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-702 <STO>

A:Cross-references: GB:AE004773; GB:AE004091; NID:g9949672; PIDN:AAG06922.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3534

```

Query Match      25.5%; Score 63; DB 2; Length 702;
Best Local Similarity 45.2%; Pred. No. 9;
Matches 14; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY      16  GNFERISGDLKQTQ--IDQVESTAGSLQGQWR 44
      || : : | | | | | | | | | |
Db      325  GNLDRVGGALCTSPALDLVASTSGGHFGRWR 355

```

RESULT 5  
A25145  
keratin, 47K type I cytoskeletal, larval - African clawed frog  
N; Alternate names: XK81

C; Species: *Xenopus laevis* (African clawed frog)  
C; Date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 13-Aug-1999  
C; Accession: A251145  
R; Jones, E.; Sargent, T. D.; Dawid, I. B.  
Proc. Natl. Acad. Sci. U.S.A. 82, 5413-5417, 1985  
A; Reference number: A251145; MUID:85270534  
A; Accession: A251145  
A; Molecule type: mRNA  
A; Residues: 1-429 <ON>  
A; Cross-references: JBN:M11940; NID:Q214565; PIDN:AAA49894.1; PID:Q214566  
C; Superfamily: cytoskeletal keratin  
C; Keywords: coiled coil

```

Query Match: 25.1%; Score 62; DB 2; Length 429;
Best Local Similarity 30.0%; Pred. No. 6.8;
Matches 12; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 3 EMKTDAAATLQGEAGNFRISGDLTKTIDQVESTAGSLQQ 42
    : : : : : : : : : : : : : : : : : : : : :
Db 303 ELQSALMKOSVGNELNQLQGFYSQQQQIQNTIGSLQQ 342

```

RESULT 6  
T02742  
probable ligand-gated ion channel protein [imported] - Arabidopsis thaliana  
N:Alternate names: hypothetical protein T914.20  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 16-Feb-2001  
C:Accession: T02742; G84692  
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;  
submitted to the EMBL Data Library, August 1998  
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.  
A:Reference number: Z14710  
A:Accession: T02742  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-934 <RD>  
A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3482941  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
Mason, T.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,  
E.; Dierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: G84692  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-934 <STO>  
A:Cross-references: GB:AE002093; NID:g3482941; PID:AA033239.1; GSPDB:GN00139

A; map position: 2  
A; Introns: 69/1; 514/3; 617/1; 748/3

Query Match	24.5%	Score 60.5;	DB 2;	Length 934;
Best Local Similarity	29.2%;	Pred. No. 25;		
Matches 19:	Conservative	8;	Mismatches 17;	Indels 21;
				Gaps 2;

```

QY      5 KTDAAATLQGE-----AGNFERISGDLKTQIDQVESTAGS---LQGW 43
      |||      ||| ||| ||| :: : || : || : || : || : || : ||
Db      330 KTNLGTGVSRYGSPSLKALSNVFENGLAGEFELINGOLESSVFDVNIIGSPERIIGLW 389

```

QY 44 RGAAG 48

Db 390 RPSNG 394

RESULT 7

RESOL  
T06388  
alpha-galactosidase (EC 3.2.1.22) - soybean



C:Species: Glycine max (soybean)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Jun-1999  
C:Accession: T06388  
R:Davis, M.O.; Walker, J.C.; Smith, D.  
Submitted to the EMBL Data Library, August 1994  
A:Description: Cloning and expression of a soybean alpha galactosidase gene.  
A:Reference number: Z15645  
A:Accession: T06388

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: mRNA

A:Residues: 1-422 <DAV>

A:Cross-references: EMBL:U12926; NID:927575; PIDN:AAA73963.1; PID:g927575

A:Experimental source: strain williams

C:Function:

A:Description: catalyzes hydrolysis of melibiose into galactose and glucose

C:Superfamily: alpha-galactosidase

C:Keywords: glycosidase; hydrolase

Query Match 24.1%; Score 59.5; DB 2; Length 422;  
Best Local Similarity 35.7%; Pred. No. 13;  
Matches 15; Conservative 7; Mismatches 19; Indels 1; Gaps 1;

QY 7 DAATLQGEAGNFRISGDLKTQIDQVESTAGSLQGOWRGAAG 48

DB 227 DPATWAKSVGNWRVTGDIQDKWDSMISRA-DLNDKWASVAG 267

RESULT 8

G70367

hypothetical protein aq\_771 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 08-Sep-2000

C:Accession: G70367

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: G70367

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-119 <AQF>

A:Cross-references: GB:AE000707; NID:g2983342; PIDN:AAC06937.1; PID:g2983349; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: aq\_771

C:Superfamily: Aquifex aeolicus hypothetical protein aq\_771

Query Match 23.5%; Score 58; DB 2; Length 119;  
Best Local Similarity 26.3%; Pred. No. 4.8;  
Matches 10; Conservative 13; Mismatches 15; Indels 0; Gaps 0;

QY 3 EMKTDATLQGEAGNFRISGDLKTQIDQVESTAGSLQ 40

DB 82 EMHREITLGNKPDESEYVEVKAEDIKKQQAANVE 119

RESULT 9

T10919

3C3.10 protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999

C:Accession: T10919

R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

Submitted to the EMBL Data Library, August 1998

A:Reference number: Z17215

A:Accession: T10919

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-124 <PAR>

A:Cross-references: EMBL:AL031231; NID:e1315070; PID:e1315080

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: 3C3.10

Query Match 23.5%; Score 58; DB 2; Length 124;  
Best Local Similarity 24.6%; Pred. No. 5;  
Matches 14; Conservative 9; Mismatches 20; Indels 14; Gaps 1;

QY 6 TDAATLQGE-----AGNFRISGDLKTQIDQVESTAGSLQGOWRGAAG 48

DB 5 TSASEGQETMSVDYSDQDLTDLADKIRTFHTDVSRRVTSNAVVDMIQGGWGAAG 61

RESULT 10

T03373

probable G-box binding factor 8 - rice

C:Species: Oryza sativa (rice)

C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000

C:Accession: T03373

R:Nakagawa, H.; Ohmiya, K.; Hattori, T.

Plant J. 9, 217-227, 1996

A:Title: A rice bZIP protein, designated OSB28, is rapidly induced by abscisic acid.

A:Reference number: Z14906; MUID:96417817

A:Accession: T03373

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: mRNA

A:Residues: 1-360 <NAK>

A:Cross-references: EMBL:U42208; NID:g1147631; PIDN:AAB40291.1; PID:g1147632

A:Experimental source: cv. Nipponbare

C:Superfamily: fava bean G-box-binding protein; fos/jun DNA-binding domain homology

Query Match

Best Local Similarity 23.5%; Score 58; DB 2; Length 360;

Matches 16; Conservative 5; Mismatches 14; Indels 8; Gaps 1;

QY 7 DAATLQGEAGNFRISGDLKTQIDQVESTAGSLQGOWRGAAGT 49

DB 122 DGATVTGSGNSEKTSQDC-----SLGSSDGNQKASGT 156

RESULT 11

C86181

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C86181

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86181

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1001 <STO>

A:Cross-references: GB:AE005172; NID:g7211984; PIDN:AAF40455.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 23.5%; Score 58; DB 2; Length 1001;

Best Local Similarity 35.9%; Pred. No. 55;

Matches 14; Conservative 7; Mismatches 18; Indels 0; Gaps 0;

QY 2 AEMKTAATLQGEAGNFRISGDLKTQIDQVESTAGSLQ 40

RESULT	14
--------	----

Search completed: July 5, 2001, 11:48:31  
Job time: 370 sec

Search completed: July 5, 2001, 11:48:31  
Job time: 370 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:51:41 ; Search time 41.8 Seconds  
(without alignments)  
40.156 Million cell updates/sec

Title: US-09-462-480-6

Perfect score: 247

Sequence: 1 MAEMKTDAAITLQGEAGNFER.....DOVESTAGSLQGWGGAAGT 49

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	25.1	429	1 K1C1_XENLA	P08777 xenopus lae
2	59.5	24.1	576	1 PEX5_PICPA	P33292 pichia past
3	57.5	23.3	906	1 PRO1_SCHPO	Q12381 schizosacch
4	56	22.7	381	1 GVPC_HALME	Q02228 halobacteri
5	56	22.7	890	1 ACOL_ECOLI	P25516 escherichia
6	55.5	22.5	378	1 AGAL_COFAR	Q42656 coffea arab
7	55.5	22.5	575	1 FLA2_CAMJE	P22251 campylobact
8	55.5	22.5	576	1 NAE2_DROME	Q9x0y0 thermotoga
9	55	22.3	975	1 K1NH_DROME	P17210 drosophila
10	55	22.3	1679	1 Y109_YEAST	P40457 saccharomyc
11	55	22.3	5179	1 MUC2_HUMAN	Q02817 homo sapien
12	54.5	22.1	575	1 FLB2_CAMJE	P22252 campylobact
13	54.5	22.1	920	1 MML7_MYCTU	P96289 mycobacteri
14	54	21.9	445	1 Y1EG_ECOLI	P31466 escherichia
15	53.5	21.7	278	1 T2D7_DROME	Q27272 drosophila
16	53	21.5	996	1 PPOL_SARPE	Q11208 sarcophaga
17	53	21.5	1019	1 POL_SIVS4	P12502 simian immu
18	53	21.5	1022	1 POL_SIVSP	P19505 simian immu
19	52.5	21.3	488	1 IMDH_HAEIN	P44334 haemophilus
20	52.5	21.3	523	1 YMR9_YEAST	Q05040 saccharomyc
21	52.5	21.3	715	1 CLPB_MYCPN	P75247 mycoplasma
22	52.5	21.3	2203	1 CCAD_RAT	P27732 rattus norv
23	52	21.1	319	1 STRN_STRGR	P29784 streptomyce
24	52	21.1	563	1 K2CA_HUMAN	P02538 homo sapien
25	52	21.1	563	1 K2CC_HUMAN	P48666 homo sapien
26	51.5	20.9	111	1 FLIE_RHIME	Q52945 rhizobium m
27	51.5	20.9	270	1 FWDC_METWO	Q74031 methanobact
28	51.5	20.9	571	1 FLA1_CAMJE	P56963 campylobact
29	51.5	20.9	571	1 FLB1_CAMJE	P56964 campylobact
30	51.5	20.9	3210	1 CENF_HUMAN	P49454 homo sapien
31	51	20.6	420	1 TRBF_AGRTU	P54914 agrobacteri
32	51	20.6	388	1 MRP4_STRPY	P30141 streptococ
33	51	20.6	574	1 FLA3_CAMJE	Q46113 campylobact

ALIGNMENTS

RESULT	ID	K1C1_XENLA	STANDARD;	PRT;	429 AA.
1	AC	P08777;			
DT	01-NOV-1988	(Rel. 09, Created)			
DT	01-NOV-1988	(Rel. 09, Last sequence update)			
DT	01-JUN-1994	(Rel. 29, Last annotation update)			
DE	KERATIN, TYPE I CYTOSKELETAL 47 KDA.				
GN	XK81A1.				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;				
OC	Xenopodinae; Xenopus.				
OX	NCBI_TaxID=8355;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=87057649; PubMed=2430981;				
RA	Miyatani S., Winkles J.A., Sargent T.D., Dawid I.B.;				
RT	"Stage-specific keratins in xenopus laevis embryos and tadpoles: the				
RT	XK81 gene family.";				
RL	J. Cell Biol. 103:1957-1965(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=85270534; PubMed=2410923;				
RA	Jonas E., Sargent T.D., Dawid I.B.;				
RT	"Epidermal keratin gene expressed in embryos of Xenopus laevis.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 82:5413-5417(1985).				
CC	-!- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.				
CC	-!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND				
CC	MICROFILAMENT KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)				
CC	(40-55 AND 56-70 KILODALTONS, RESPECTIVELY).				
CC	-!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; X04804; CAA28496.1; -				
DR	EMBL; X04668; CAA28374.1; -				
DR	EMBL; M11940; AAA49894.1; -				
DR	PIR; A25145; A25145.				
DR	InterPro; IPR001664; -				
DR	InterPro; IPR002957; -				
DR	Protein; P00038; filament; 1.				
DR	PRINTS; PR01248; TYPEIKERATIN.				
DR	PROSITE; PS00226; IF; 1.				
KW	Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin.				
FT	DOMAIN 1 69				
FT	DOMAIN 70 381				
FT	DOMAIN 382 429				
FT	DOMAIN 70 105				
FT	DOMAIN 106 127				
FT	LINKER 1.				

Q46114 campylobact  
P50176 r poly-beta  
P79955 xenopus lae  
P03815 escherichia  
P05659 acanthamoeb  
P23467 homo sapien  
P28475 archaeoglob  
P74330 synechocyst  
P14251 sendai viru  
Q01401 oryza sativ  
P29693 xenopus lae  
P43800 haemophilus

1 FLB3\_CAMJE  
1 PHBC\_RHIME  
1 CTK2\_XENLA  
1 CLPB\_ECOLI  
1 MYSN\_ACACA  
1 PTPB\_HUMAN  
1 YH99\_ARCFU  
1 SYC\_SYNY3  
1 RRPB\_SEND6  
1 GLGB\_ORYSA  
1 EFID\_XENLA  
1 GLPB\_HAEIN

51 20.6 574  
51 20.6 611  
51 20.6 643  
51 20.6 857  
51 20.6 1509  
51 20.6 1997  
50.5 20.4 226  
50.5 20.4 483  
50.5 20.4 568  
50.5 20.4 820  
50 20.2 265  
50 20.2 432

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Z19592; CAA79640.1; -  
EMBL; U59222; AAB40613.1; -

--!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
--!- SIMILARITY: TO SOME OTHER PRE-MRNA SPLICING FACTORS.  
--!- SIMILARITY: CONTAINS 19 TPR REPEATS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)) or send an email to [license@sib-stb.ch](mailto:license@sib-stb.ch)







CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AE001780; AAD36328.1; -.
DR	TIGR; TM1253; -.
DR	Pfam; PF00795; CN_hydrolase; 1.
DR	PROSITE; PS00920; NITRIL_CHT_1; UNKNOWN_1.
DR	Ligase; NAD; ATP-binding.
FT	DOMAIN 292 576
FT	NP_BIND 321 328
FT	ACT_SITE 323 323
FT	SEQUENCE 576 AA; 64919 MW; 208FDC77964C957F CRC64;
CC	-----
CC	-----

```

Query Match      22.5%; Score 55.5; DB 1; Length 576;
Best Local Similarity 37.8%; Pred. No. 22;
Matches 17; Conservative 4; Mismatches 19; Indels 5; Gaps 2;

QY      2  AENKTDAAIT-GGEAGNFERISGDLTKTQIDOVESTAGSLQGWGRG 45
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      432  SENATGYATLYGDMAGGFAVIKDVYKTDVYRI---GRWYNSWRG 472

```

KINL_DROME	KINL_DROME	STANDARD;	PRT;	975 AA.
ID	PI7210;			
AC	01-AUG-1990	(Rel. 15, Created)		
DT	01-AUG-1990	(Rel. 15, Last sequence update)		
DT	15-JUL-1998	(Rel. 36, Last annotation update)		
DE	KINESIN HEAVY CHAIN.			
GN	KHC OR KIN.			

Eukaryota; Metazoa; Arthropoda; Insecta; Muscomorpha;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=89168428; PubMed=2523352;  
 RA Yang J.T., Laymon R.A., Goldstein L.S.B.;  
 RT "A three-domain structure of kinesin heavy chain revealed by DNA  
 RL sequence and microtubule binding analyses.";  
 RN Cell 56:879-889(1989).  
 [2]  
 RN MUTAGENESIS.  
 RX MEDLINE=93030741; PubMed=1384131;  
 RA Gho M., McDonald K., Ganetzky B., Saxton W.M.;  
 RT "Effects of Kinesin mutations on neuronal functions.";  
 RN Science 258:313-316(1992).  
 RL -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING  
 CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.  
 CC -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT  
 CC CHAINS.  
 CC -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-  
 CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF  
 CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL  
 CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN  
 CC DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH  
 CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),  
 CC VESICLES AND MEMBRANOUS ORGANELLES.  
 CC -!- MISCELLANEOUS: MUTANT FLIES DISPLAY IMPAIRED ACTION POTENTIAL  
 CC PROPAGATION AND NEUROTRANSMITTER RELEASE AT NEUROMUSCULAR  
 CC JUNCTIONS, BUT ARE STILL CAPABLE OF TRANSPORTING CERTAIN  
 CC MEMBRANES, INCLUDING SYNAPTIC VESICLES, TO THE NERVE TERMINAL.  
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN  
 CC SUBFAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; M24441; AAA28652.1; --  
DR PIR; A31497; A31497.  
DR HSP; P56536; 2KIN.  
DR FlyBase; FBgn0001308; Khc.  
DR InterPro; IPR001752; --  
DR Pfam; PF00225; Kinesin; 1.  
DR PRINTS; PR00380; KINESINHEAVY.  
DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.  
KW Motor protein; ATP-binding; Coiled coil.  
FT DOMAIN 1 334 MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).  
FT DOMAIN 335 931 COILED COIL.  
FT DOMAIN 932 975 GLOBULAR.  
FT DOMAIN 180 321 MICROTUBULE-BINDING.  
FT NP\_BIND 92 99 ATP (BY SIMILARITY).  
SQ SEQUENCE 975 AA; 110429 MW; 9966C6C35BA74FD6 CRC64;

Query Match 22.3%; Score 55; DB 1; Length 975;  
Best Local Similarity 29.7%; Pred. No. 46;  
Matches 11; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGOBAGNFERISGDLKTQIDQVESTAG 37  
Db 618 ISKMKTEAKNIAQRCSNETQADSNKISEYKDLG 654

RESULT 10  
Y109\_YEAST STANDARD; PRT; 1679 AA.  
AC P40457;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DE 01-FEB-1995 (Rel. 31, Last annotation update)  
DE HYPOTHETICAL 195.1 KDA PROTEIN IN DNA43-UBI1 INTERGENIC REGION.  
GN YIL149C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
RA Walsh S.V., Whitehead S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; Z38059; CAA86129.1; --  
DR PIR; S48385; S48385.  
DR SGD; S0001411; MLP2.  
KW Hypothetical protein.  
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;

Query Match 22.3%; Score 55; DB 1; Length 1679;  
Best Local Similarity 29.3%; Pred. No. 83;  
Matches 12; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 9 ATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWRAAGT 49  
Db 778 AKIKLENNLERLSDLQSKIQEIESIRSCDSQLKWAQNT 818

RESULT 11  
MUC2\_HUMAN STANDARD; PRT; 5179 AA.  
ID MUC2\_HUMAN  
AC Q02817; Q14878;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).  
GN MUC2 OR SMUC.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Intestine;  
RX MEDLINE=94132002; PubMed=8300571;  
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;  
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.  
RT Identification of the amino terminus and overall sequence similarity  
RT to prepro-von Willebrand factor.";  
RL J. Biol. Chem. 269:2440-2446(1994).  
RN [2]  
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
RC TISSUE=Colon;  
RX MEDLINE=93016075; PubMed=1400449;  
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,  
RA Kim Y.S.;  
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
RT both upstream and downstream of its central repetitive region.";  
RL J. Biol. Chem. 267:21375-21383(1992).  
RN [3]  
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.  
RX MEDLINE=91358717; PubMed=1885763;  
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
RA Petersen G.M., Kim Y.S.;  
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
RT and polymorphism.";  
RL J. Clin. Invest. 88:1005-1013(1991).  
CC -1- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND  
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A  
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS  
CC AGENTS AT MUCOSAL SURFACES.  
CC -1- SUBUNIT: MULTIMERIC.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,  
CC BRONCHUS, CERVIX AND GALL BLADDER.  
CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND  
CC VARIES AMONG DIFFERENT ALLELES.  
CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT  
CC OF SILKWORM HEMOCYTIN.  
CC -1- SIMILARITY: CONTAINS 2 WVFC DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; L21998; AAB95295.1; --  
DR EMBL; M74027; AAA59875.1; --  
DR EMBL; M94131; AAA59163.1; --

DR EMBL; M94132; AAA59164.1; -  
 DR MIM; 158370; -  
 DR InterPro; IPR000359; -  
 DR InterPro; IPR000561; -  
 DR InterPro; IPR001007; -  
 DR InterPro; IPR001840; -  
 DR InterPro; IPR002400; -  
 DR Pfam; PF00007; Cys\_knot; 1.  
 DR Pfam; PF00094; vwd; 4.  
 DR PRINTS; PR00438; GFCYSKNOT.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01185; CTCK\_1; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR PROSITE; PS01208; WFEC; 2.  
 KW Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 5179  
 FT DOMAIN 1401 1747  
 FT REPEAT 1401 1416 1.  
 FT REPEAT 1417 1432 2.  
 FT REPEAT 1433 1448 3.  
 FT REPEAT 1449 1464 4.  
 FT REPEAT 1465 1471 5.  
 FT REPEAT 1472 1478 6.  
 FT REPEAT 1479 1494 7A.  
 FT REPEAT 1495 1517 7B.  
 FT REPEAT 1518 1533 8A.  
 FT REPEAT 1534 1556 8B.  
 FT REPEAT 1557 1572 9A.  
 FT REPEAT 1573 1596 9B.  
 FT REPEAT 1597 1612 10A.  
 FT REPEAT 1613 1635 10B.  
 FT REPEAT 1636 1651 11A.  
 FT REPEAT 1652 1675 11B.  
 FT REPEAT 1676 1683 12.  
 FT REPEAT 1684 1699 13.  
 FT REPEAT 1700 1715 14.  
 FT REPEAT 1716 1731 15.  
 FT REPEAT 1732 1747 16.  
 FT DOMAIN 4815 4886 WFEC 1.  
 FT DOMAIN 4924 4991 WFEC 2.  
 FT DOMAIN 5075 5160 CTCK.  
 FT DISULFID 5075 5122 BY SIMILARITY.  
 FT DISULFID 5089 5136 BY SIMILARITY.  
 FT DISULFID 5098 5152 BY SIMILARITY.  
 FT DISULFID 5102 5154 BY SIMILARITY.  
 FT DISULFID ? 5159 BY SIMILARITY.  
 FT CARBOHYD 163 183 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 423 433 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 770 770 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 894 894 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CONFLICT 1351 1351 H -> L (IN REF. 3).  
 FT CONFLICT 1412 1412 T -> S (IN REF. 3).  
 FT CONFLICT 1449 1449 L -> P (IN REF. 3).  
 FT CONFLICT 1504 1504 M -> T (IN REF. 3).  
 FT CONFLICT 4192 4192 G -> S (IN REF. 2).  
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;  
 Query Match 22.3%; Score 55; DB 1; Length 5179;  
 Best Local Similarity 35.7%; Pred. No. 2.8e+02;  
 Matches 15; Conservative 4; Mismatches 19; Indels 4; Gaps 2;  
 QY 7 DAATLQGE---ACNFERISG-DLKTQIDQVESTAGSLQGW 44  
 DB 517 DQASQGVQVGLCGFNGLGDDFTAGSLVEATGAGFANTWK 558  
 RESULT 12  
 FLB2\_CAMJE STANDARD; PRT; 575 AA.  
 AC P22252;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE FLAGELLIN B.  
 GN FLAB.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=81116;  
 RX MEDLINE=91009243; PubMed=2211662;  
 RA Nuijten P.J., van Asten F.J., Gaastera W., van der Zeijst B.A.;  
 RT "Structural and functional analysis of two Campylobacter jejuni  
 flagellin genes";  
 RL J. Biol. Chem. 265:17798-17804(1990).  
 CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
 FORM THE FILAMENTS OF BACTERIAL FLAGELLA.  
 CC -!- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collabora-  
 tion between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; J05635; AAA23025.1; -  
 CC PIR; B39228; B39228.  
 CC InterPro; IPR001029; -  
 CC InterPro; IPR001492; -  
 CC Pfam; PF00700; Flagellin\_C; 1.  
 CC Pfam; PF00669; Flagellin\_N; 1.  
 CC PRINTS; PR00207; FLAGELLIN.  
 KW Flagella.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 SQ SEQUENCE 575 AA; 59728 MW; D0531AF308A7BF1D CRC64;  
 Query Match 22.1%; Score 54.5; DB 1; Length 575;  
 Best Local Similarity 28.0%; Pred. No. 30;  
 Matches 14; Conservative 7; Mismatches 20; Indels 9; Gaps 1;  
 QY 2 AEKMTDAATLQGEAGNFERISGDLK-----TQIDQVESTAGSLQGW 42  
 DB 517 DQASQGVQVGLCGFNGLGDDFTAGSLVEATGAGFANTWK 558



Search completed: July 5, 2001, 11:51:42  
Job time: 501 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:50:46 ; Search time 123.78 seconds  
(without alignments)  
52.375 Million cell updates/sec

Title: US-09-462-480-6  
Perfect score: 247  
Sequence: 1 MAEKTDAATLQGEAGNFER.....DQVESTAGSLQGWKRAAGT 49

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTRMBL\_16.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_unclassified.\*
- 13: sp\_vertebrate.\*
- 14: sp\_virus.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241	97.6	100	2 069739	069739 mycobacteri
2	92	37.2	100	2 033084	033084 mycobacteri
3	63	25.5	103	2 005440	005440 mycobacteri
4	63	25.5	702	2 09H176	09H176 pseudomonas
5	62	25.1	2954	14 096898	096898 hepatitis g
6	60.5	24.5	934	10 081080	081080 arabidopsis
7	59.5	24.1	422	10 039811	039811 glycine max
8	59	23.9	203	5 09VD64	09VD64 drosophila
9	58.5	23.7	944	4 09UP82	09UP82 homo sapien
10	58.5	23.7	1045	4 076037	076037 homo sapien
11	58.5	23.7	1107	4 09H4D6	09H4D6 homo sapien
12	58.5	23.7	1157	4 09G2M2	09G2M2 homo sapien
13	58.5	23.7	1187	4 09G2V6	09G2V6 homo sapien
14	58.5	23.7	1199	4 09H487	09H487 homo sapien
15	58.5	23.7	1201	4 09H486	09H486 homo sapien
16	58.5	23.7	1214	4 09H485	09H485 homo sapien
17	58.5	23.7	1215	4 09H484	09H484 homo sapien
18	58.5	23.7	1256	4 09H483	09H483 homo sapien
19	58.5	23.7	1827	4 09H482	09H482 homo sapien

20	58.5	23.7	2117	4 09H4D8	09H4D8 homo sapien
21	58.5	23.7	2167	4 09H481	09H481 homo sapien
22	58.5	23.7	2169	4 09NV09	09NV09 homo sapien
23	58	23.5	119	2 066970	066970 aquifex aeo
24	58	23.5	124	2 086643	086643 streptomyce
25	58	23.5	360	10 040645	040645 oryza sativ
26	58	23.5	1001	10 09MAT0	09MAT0 arabidopsis
27	58	23.5	1082	2 09JXM7	09JXM7 neisseria m
28	58	23.5	1084	2 09JPL3	09JPL3 neisseria m
29	58	23.5	1313	10 09XIP6	09XIP6 arabidopsis
30	58	23.5	1829	2 086808	086808 streptomyce
31	57	23.1	248	5 090952	090952 caenorhabdi
32	57	23.1	333	2 09KC26	09KC26 bacillus ha
33	57	23.1	3906	2 09RNB2	09RNB2 microcystis
34	57	23.1	3906	2 09FDU1	09FDU1 microcystis
35	56.5	22.9	290	2 09L216	09L216 streptomyce
36	56.5	22.9	425	10 041100	041100 phaseolus v
37	56.5	22.9	1186	2 09ZBQ2	09ZBQ2 streptomyce
38	56	22.7	417	2 055964	055964 synchocys
39	56	22.7	458	3 09UW11	09UW11 piromyces r
40	56	22.7	1061	9 064335	064335 bacterioph
41	55.5	22.5	574	2 09RPY6	09RPY6 campylobact
42	55.5	22.5	576	2 085179	085179 campylobact
43	55.5	22.5	576	2 09RF26	09RF26 campylobact
44	55.5	22.5	576	2 09R953	09R953 campylobact
45	55.5	22.5	576	2 09R950	09R950 campylobact

## ALIGNMENTS

RESULT	1				
069739					
ID	069739	PRELIMINARY;	PRT;	100 AA.	
AC	069739;				
DT	01-AUG-1998 (TREMBLrel. 07, Created)				
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)				
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)				
DE	HYPOTHETICAL 10.8 KDA PROTEIN.				
GN	LHP.OR RV3874 OR MTV027.09.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,				
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Sulston J.E.,				
RA	Taylor K., Whitehead S., Barrell B.G.;				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RA	Berthet F.-X., Birk Rasmussen P., Andersen P., Gicquel B.;				
RT	"Promoter analysis of the M. tuberculosis orfIC gene encoding the				
RT	early secreted antigenic target 6 kDa (ESAT-6)."				
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AL022120; CAA17966.1; -				
DR	EMBL; AF004671; AAC83445.1; -				
DR	TubercuList; Rv3874; -				
KW	Hypothetical protein.				
SQ	SEQUENCE 100 AA; 10794 MW; 285F4FC96F55D194 CRC64;				





RA Simons J.N., Pilot-Matias T.J., Barry T.P., Dawson G.J., Desai S.M.,  
RA Schlauder G.G., Muerhoff A.S., Walker J.C., Buijk S.L., Chalmers M.L.,  
RA van Sant C.L., Mushawar I.K.;  
RT "Identification of two flavivirus-like genomes in the GB hepatitis  
RT agent.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:3401-3405(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96323131; PubMed=8709237;  
RA Simons J.N., Desai S.M., Schultz D.E., Lemon S.M., Mushawar I.K.;  
RT "Translation Initiation in GB viruses A and C: evidence for internal  
RT ribosome entry and implications for genome organization.";  
RL J. Virol. 70:6126-6135(1996).  
CC 1- SIMILARITY: TO HEPATITIS C VIRUS RNA DEPENDENT RNA POLYMERASE.  
DR EMBL: U22303; AAC55983.1; -;  
DR HSSP: P27958; 1HEI.  
DR MEROPS: S29.002; -;  
DR InterPro: IPR000255; -;  
DR InterPro: IPR000745; -;  
DR InterPro: IPR001410; -;  
DR InterPro: IPR001490; -;  
DR InterPro: IPR002166; -;  
DR InterPro: IPR002518; -;  
DR InterPro: IPR002868; -;  
DR Pfam: PF00998; HCV\_RGRP; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR PROSITE: PS00012; PHOSPHOPANTHETINE; UNKNOWN\_1.  
DR SMART: SM00487; DEXDC; 1.  
KW Nonstructural protein; RNA-directed RNA polymerase.  
SQ SEQUENCE 2954 AA; 318835 MW; 4A0C35FC2ED283B1 CRC64;

Query Match 25.1%; Score 62; DB 14; Length 2954;  
Best Local Similarity 34.8%; Pred. No. 1.4e+02;  
Matches 16; Conservative 4; Mismatches 20; Indels 6; Gaps 1;  
QY 7 DAATLGOEAGNFE-----RISGLKTDQIDQVESTAGSLQGWGA 46  
DB 2357 DGHQFDERGPIEVPTICEVIGIRLQCDQIBETPTYSYIWSGA 2402

RESULT 6  
ID 081080 PRELIMINARY; PRT; 934 AA.  
AC 081080;  
DT 01-NOV-1998 (TEMBLrel. 08, Created)  
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)  
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
DE PUTATIVE LIGAND-GATED IONIC CHANNEL.  
GN T914.20.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,  
RA Shen M., Rönning C.M., Fraser C.M., Somerville C.R., Venter J.C.;  
RT "Arabidopsis thaliana chromosome II BAC T914 genomic sequence.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC005315; AC33339.1; -;  
DR InterPro: IPR001311; -;  
DR InterPro: IPR001320; -;  
DR InterPro: IPR001395; -;  
DR InterPro: IPR001828; -;  
DR Pfam: PF00060; lig\_chan; 1.  
DR Pfam: PF01094; ANF\_receptor; 1.  
DR PROSITE: PS00063; ALDOKETO\_REDUCTASE\_3; UNKNOWN\_1.

DR SMART: SM00079; PPPE; 1.  
SQ SEQUENCE 934 AA; 105052 MW; 6A325C48CF4E2B4E CRC64;

Query Match 24.5%; Score 60.5; DB 10; Length 934;  
Best Local Similarity 29.2%; Pred. No. 56;  
Matches 19; Conservative 8; Mismatches 17; Indels 21; Gaps 2;  
QY 5 KTDAAATLGOE-----AGNFERISGLKTDQIDQVESTAGS---LOGQW 43  
DB 330 KTLNGLTGVSRYPGSLKALSNVRFNGLAGEFELINGQLQESSVDFVINIIGSEERIGLW 389  
QY 44 RGAAG 48  
DB 390 RPSNG 394

RESULT 7  
ID Q39811 PRELIMINARY; PRT; 422 AA.  
AC Q39811;  
DT 01-NOV-1996 (TEMBLrel. 01, Created)  
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)  
DE ALPHA GALACTOSIDASE.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;  
OC Fabales; Fabaceae; Papilionoideae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WILLIAMS;  
RA Davis M.O., Walker J.C., Smith D.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U12926; AAA73963.1; -;  
DR Mendel; 8332; Glyma; 1501.8332.  
DR InterPro: IPR000111; -;  
DR InterPro: IPR002241; -;  
DR InterPro: IPR002287; -;  
DR Pfam: PF02065; Melibiose; 2.  
DR PRINTS: PR00740; GLHYDRLASE27.  
DR PRODOM: PD002572; -; 1.  
DR PROSITE: PS00512; ALPHA\_GALACTOSIDASE; 1.  
SQ SEQUENCE 422 AA; 46395 MW; 9281B25EAD6940A0 CRC64;

Query Match 24.1%; Score 59.5; DB 10; Length 422;  
Best Local Similarity 35.7%; Pred. No. 30;  
Matches 15; Conservative 7; Mismatches 19; Indels 1; Gaps 1;  
QY 7 DAATLGOEAGNFERISGLKTDQIDQVESTAGSLQGWGAAG 48  
DB 227 DPATWAKSVGNWRTTGDQDKWDSMISRA-DLNDKRWASTAG 267

RESULT 8  
ID Q9VD64 PRELIMINARY; PRT; 203 AA.  
AC Q9VD64;  
DT 01-MAY-2000 (TEMBLrel. 13, Created)  
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
DE CG5650 PROTEIN.  
GN CG5650.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;





DR EMBL; AJ400853; CAC14138.1; -.  
SQ SEQUENCE 1201 AA; 123504 MW; 8C635BEAB245C5DC CRC64;

```

Query Match      23.7%; Score 58.5; DB 4; Length 1201;
Best Local Similarity 39.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 5; Mismatches 19; Indels 1; Gaps 1;

```

QY 9 ATLQEQAGNFERISGLTKTIDQVESTAGSLQGWRAAGT 49  
 |||| | : | | : | | | : | | | :  
 Db 727 ATLGPSTGSLTKGAL-TLANSVVSTPGPEGOWTSAS 766

Search completed: July 5, 2001, 11:50:47  
Job time: 470 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2002, 21:27:53 ; Search time 58.34 seconds  
(without alignments)  
190.390 Million cell updates/sec

Title: US-09-462-480-5  
Perfect score: 492  
Sequence: 1 MAEMKTDATLQEGAGNER.....VOYSRADEEQQALSSQMGF 100

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	492	100.0	100	20	AAV03705 M. tuberculosis LH
2	486	98.8	100	19	AAW81706 M. tuberculosis im
3	486	98.8	100	19	AAW64339 Mycobacterium tube
4	486	98.8	100	20	AAV39136 M. tuberculosis an
5	486	98.8	100	20	AAV38993 M. tuberculosis re
6	486	98.8	100	22	AAV35218 M. tuberculosis RV3
7	486	98.8	100	22	AAV19845 Mycobacterium tube
8	481	97.8	802	19	AAW81746 M. tuberculosis fu
9	481	97.8	802	19	AAW64379 Mycobacterium anti
10	481	97.8	802	20	AAV32063 Mycobacterium tube
11	481	97.8	802	20	AAV39224 M. tuberculosis fu

12	481	97.8	802	20	AAV39176 M. tuberculosis fu
13	481	97.8	802	20	AAV39081 M. tuberculosis fus
14	481	97.8	802	20	AAV39033 M. tuberculosis fu
15	481	97.8	983	22	AAU01901 M. tuberculosis Tb
16	462	93.9	95	18	AAW32444 Mycobacterium tube
17	462	93.9	95	18	AAW32376 Mycobacterium tube
18	462	93.9	95	19	AAW81747 M. tuberculosis im
19	462	93.9	95	19	AAW64321 Mycobacterium tube
20	462	93.9	95	20	AAV32097 Mycobacterium tube
21	462	93.9	95	20	AAV39118 M. tuberculosis an
22	462	93.9	95	20	AAV38981 M. tuberculosis re
23	392	79.7	80	18	AAW32454 Mycobacterium tube
24	392	79.7	80	18	AAW32386 Mycobacterium tube
25	392	79.7	80	19	AAW81707 M. tuberculosis im
26	392	79.7	80	19	AAW64340 Mycobacterium tube
27	392	79.7	80	20	AAV39137 M. tuberculosis an
28	392	79.7	80	20	AAV38994 M. tuberculosis re
29	247	50.2	49	20	AAV03706 M. tuberculosis LH
30	205	41.7	42	20	AAV03707 M. tuberculosis LH
31	137	27.8	28	20	AAV03712 M. tuberculosis LH
32	122	24.8	28	18	AAW32460 Mycobacterium tube
33	122	24.8	28	19	AAW81698 M. tuberculosis im
34	122	24.8	28	20	AAV39128 M. tuberculosis an
35	121	24.6	27	18	AAW32458 Mycobacterium tube
36	121	24.6	27	19	AAW81696 M. tuberculosis im
37	121	24.6	27	20	AAV39126 M. tuberculosis an
38	118	24.0	27	18	AAW32457 Mycobacterium tube
39	118	24.0	27	18	AAW32459 Mycobacterium tube
40	118	24.0	27	19	AAW81695 M. tuberculosis im
41	118	24.0	27	19	AAW81697 M. tuberculosis im
42	118	24.0	27	20	AAV39125 M. tuberculosis an
43	118	24.0	27	20	AAV39127 M. tuberculosis an
44	114	23.2	28	18	AAW32455 Mycobacterium tube
45	114	23.2	28	19	AAW81693 M. tuberculosis im

ALIGNMENTS

RESULT 1  
AAV03705  
ID AAV03705 standard; Protein: 100 AA.  
XX  
AC AAV03705;  
DT 07-JUN-1999 (first entry)  
XX  
DE M. tuberculosis LHP polypeptide.  
XX  
ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
KW immune response.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO9904005-A1.  
XX  
PD 28-JAN-1999.  
XX  
PF 16-JUL-1998; 98WO-IB01091.  
XX  
PR 16-JUL-1997; 97US-0052631.  
XX  
PA (INSP ) INST PASTEUR.  
PA (STAT-) STATENS SERUM INST.  
XX  
PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;  
XX  
DR WPI: 1999-132249/11.  
XX  
N-PSDB: AAX29168, AAX29171.  
XX  
New nucleic acid containing regulator and LHP gene of Mycobacterium  
PT tuberculosis - useful in vaccines, for diagnosis, and for expression

PT of heterologous proteins

PS Claim 17; Page 64; 88pp; English.

XX The present invention is directed to a polynucleotide carrying the  
 CC regulatory expression signals of the ESAT-6 protein as well as an open  
 CC reading frame coding for an antigenic protein LHP from Mycobacterium  
 CC tuberculosis. Host cells comprising the polynucleotide are used for the  
 CC recombinant expression of the protein. The recombinant polypeptide can be  
 CC used as immunogens and vaccines, to protect against bacteria of the  
 CC M. tuberculosis complex in humans or animals (the vaccines may include  
 CC other immunogenic proteins of the bacteria or their fragments,  
 CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by  
 CC detection of specific antibodies. The regulatory region present in the  
 CC polynucleotide may be used to express almost any heterologous protein in  
 CC mycobacteria, particularly as a fusion with polyhistidine. The two  
 CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to  
 CC provide a synergistic increase in ability to induce a protective immune  
 CC response. The present sequence represents the LHP polypeptide.

XX Sequence 100 AA;

Query Match 100.0%; Score 492; DB 20; Length 100;

Best Local Similarity 100.0%; Pred. No. 4e-46;

Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAEMKTDATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWGAAGTAAQAQAAVVRFOE 60

Db 1 maemktdaatlqgeagnferisgdlktqldqvvestagslqgwrgaagtaaqaaavvrifqe 60

Qy 61 AANKQKQELDEISNIRAGVOYVSRADEEQQALSSQMGF 100

Db 61 aankqkqeldeisniragvgvysradeeqqalssqmgf 100

RESULT 2

AAW81706  
 ID AAW81706 standard; Protein; 100 AA.

XX AAW81706;

XX 27-JAN-1999 (first entry)

DE M. tuberculosis immunogenic polypeptide Tb38-IN.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX WO9816646-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18293.

XX 13-MAR-1997; 97US-0818112.

XX 11-OCT-1996; 96US-0730510.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis

XX Example 3B; Page 138-139; 230pp; English.

XX

CC This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.

XX Sequence 100 AA;

Query Match 98.8%; Score 486; DB 19; Length 100;

Best Local Similarity 99.0%; Pred. No. 1.8e-45;

Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MAEMKTDATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWGAAGTAAQAQAAVVRFOE 60

Db 1 maemktdaatlqgeagnferisgdlktqldqvvestagslqgwrgaagtaaqaaavvrifqe 60

Qy 61 AANKQKQELDEISNIRAGVOYVSRADEEQQALSSQMGF 100

Db 61 aankqkqeldeisniragvgvysradeeqqalssqmgf 100

RESULT 3

AAW64339

ID AAW64339 standard; Protein; 100 AA.

XX AAW64339;

XX 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen Tb38-IN.

XX Tuberculosis; infection; diagnosis; antigen; Tb37-FL.

XX Mycobacterium tuberculosis strain H37Rv.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

XX 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and diagnosis of tuberculosis

XX Example 3; Page 145; 250pp; English.

XX This polypeptide comprises a partial sequence of Mycobacterium  
 CC tuberculosis antigen Tb38-IN. It is encoded by genomic DNA isolated  
 CC from a M. tuberculosis strain H37Rv genomic library using a probe  
 CC derived from clone Tb38-1 (see AAV44384). The invention relates to  
 CC compositions and methods for diagnosing tuberculosis. It provides  
 CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of  
 CC a soluble M. tuberculosis antigen, or an immunogenic portion of an  
 CC M. tuberculosis antigen, as well as DNA sequences encoding such  
 CC polypeptides, recombinant expression vectors and transformed or  
 CC transfected host cells. Also claimed are methods and diagnostic  
 CC kits for detecting M. tuberculosis infection in a patient using  
 CC these polypeptides, antibodies or oligonucleotide probes and  
 CC primers, for the diagnosis of tuberculosis.

XX

QY 1 MAEMKTDATLGOEAGNFERISGDLKTOIDOVESTAGSLOGWARGAAGTAAOAAVVRFOE 60

AA  
AC AAR35218.

DT 24-APR-2001 (first entry)  
 XX M tuberculosis Rv3874 protein.  
 XX  
 DE Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c;  
 XX Rv1037c; Rv2348c; Rv2654c; Rv3020c; Rv3444c;  
 KW Rv3445c; Rv3890c; Rv3891c; Rv3904c; Rv3905c.  
 KW  
 XX Mycobacterium tuberculosis.  
 OS  
 XX WO200104151-A2.  
 PN  
 XX 18-JAN-2001.  
 XX  
 PD  
 XX 13-JUL-2000; 2000WO-DK00398.  
 PF  
 XX 13-JUL-1999; 99DK-0001020.  
 PR  
 XX 15-JUL-1999; 99US-0144011.  
 XX  
 XX (STAT-) STATENS SERUM INST.  
 PA  
 XX Andersen P, Skjot R;  
 PI  
 XX WPI; 2001-091923/10.  
 DR  
 XX New polypeptide encoded by a member of the esat-6-gene family for  
 PT immunizing against and diagnosis of tuberculosis.  
 XX  
 XX Example 2; Page 65; 80pp; English.  
 PS  
 XX The present invention provides the protein and coding sequences for  
 CC members of the esat-6 gene family from Mycobacterium tuberculosis. These  
 CC proteins include Rv0287, Rv1036c, Rv1037c, Rv2348c, Rv2653c,  
 CC Rv2654c, Rv3020c, Rv3444c, Rv3445c, Rv3890c, Rv3891c, Rv3904c and  
 CC Rv3905c. These can be used to produce vaccines against, and in the  
 CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of  
 CC the proteins of the invention.  
 XX  
 XX Sequence 100 AA;  
 SQ  
 Query Match 98.8%; Score 486; DB 22; Length 100;  
 Best Local Similarity 99.0%; Pred. No. 1.8e-45;  
 Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MAEMKTDATLGOEAGNFERISGDLKTDIDOVSTAGSLQGWGAGTAQAQAAVVRQEE 60  
 DB 1 maemktdaatlaqea gnferisgdlktidqvstagslqgwrgaagtaagaavvrqfe 60  
 QY 61 AANKOKQELDEISTNIRAGVOYISRADEEQOQALSSQMGF 100  
 DB 61 aankokqeldeistniragvqvysradeeqqalssqmgf 100  
 RESULT 7  
 AAB19845  
 ID AAB19845 standard; Protein; 100 AA.  
 XX  
 AC AAB19845;  
 XX  
 XX 05-MAR-2001 (first entry)  
 DT  
 XX Mycobacterium tuberculosis protein MTBN4.  
 DE  
 XX MTBN4; tuberculosis; BCG; vaccine; infection; diagnosis.  
 XX  
 XX Mycobacterium tuberculosis.  
 OS  
 XX WO200066157-A1.  
 PN  
 XX 09-NOV-2000.  
 PD  
 XX 04-MAY-2000; 2000WO-US12257.  
 PF

XX 04-MAY-1999; 99US-0132505.  
 PR  
 XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.  
 PA  
 XX Gennaro ML;  
 PI  
 XX WPI; 2001-007153/01.  
 DR  
 XX N-PSDB; AAA89038.  
 XX  
 XX Novel polypeptide encoded by open reading frames present in  
 PT Mycobacterium tuberculosis genome and not by the BCG strain of M.  
 PT bovis, useful as vaccine and for diagnosing tuberculosis infection  
 XX  
 XX Claim 11; Fig 1; 35pp; English.  
 PS  
 XX The present sequence is that of the Mycobacterium tuberculosis  
 CC MTBN4 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see  
 CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)  
 CC identified as being present in the genome of M. tuberculosis but  
 CC absent from the genome of the BCG strain of Mycobacterium bovis.  
 CC MTBN1-8 represent reagents that are useful in discriminating between  
 CC M. tuberculosis and BCG and, in particular, for diagnostic methods  
 CC which discriminate between exposure of a subject to M. tuberculosis  
 CC and vaccination with BCG. The invention features these MTBN  
 CC polypeptides, functional fragments of them, DNA encoding them,  
 CC vectors, transformed cells, and diagnostic, therapeutic, and  
 CC prophylactic (vaccine) methods, including genetic vaccination  
 CC methods.  
 XX  
 XX Sequence 100 AA;  
 SQ  
 Query Match 98.8%; Score 486; DB 22; Length 100;  
 Best Local Similarity 99.0%; Pred. No. 1.8e-45;  
 Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MAEMKTDATLGOEAGNFERISGDLKTDIDOVSTAGSLQGWGAGTAQAQAAVVRQEE 60  
 DB 1 maemktdaatlaqea gnferisgdlktidqvstagslqgwrgaagtaagaavvrqfe 60  
 QY 61 AANKOKQELDEISTNIRAGVOYISRADEEQOQALSSQMGF 100  
 DB 61 aankokqeldeistniragvqvysradeeqqalssqmgf 100  
 RESULT 8  
 AAW81746  
 ID AAW81746 standard; Protein; 802 AA.  
 XX  
 AC AAW81746;  
 XX  
 XX 27-JAN-1999 (first entry)  
 DT  
 XX M. tuberculosis fusion protein TbF-2.  
 DE  
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 XX  
 XX Synthetic.  
 OS  
 XX Mycobacterium tuberculosis.  
 OS  
 XX WO9816646-A2.  
 PN  
 XX 23-APR-1998.  
 PD  
 XX 07-OCT-1997; 97WO-US18293.  
 PF  
 XX 13-MAR-1997; 97US-0818112.  
 PR  
 XX 11-OCT-1996; 96US-0730510.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA



PS Example 7; page 223-226; 250pp; English.

CC AAY39225 are used in the exemplification of the present invention.

XX

SQ Sequence 802 AA;

Query Match 97.8%; Score 481; DB 20; Length 802;  
Best Local Similarity 99.0%; Pred. No. 8.1e-44;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps

Qy 2 AEMKTDATLQGEAGNFERISGDLKTIDQVESTAGSLQGWRAAGTAAQAAVVRFOEA 61  
|||||

Db 425 aemktdaatlaeagneferisgdlktidqvestagslqgwgaaagtaaaqaavvrfoea 484  
|||||

Qy 62 ANKKQKQELDEISTNIRAGVQVYSRADEEQQALSSQMGF 100  
|||||

Db 485 ankqkqeldeistniragvgysradeeqqggaissqmgf 523  
|||||

RESULT 12

AAY39176

ID AAY39176 standard; Protein; 802 AA.

XX

AC AAY39176;

XX

DT 05-NOV-1999 (first entry)

XX

DE M. tuberculosis fusion protein TbF-2 amino acid sequence.

XX

KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
immunotherapy; diagnosis; immunisation; vaccine; infection;  
immune response; skin test.

KW

KW

KW

XX

OS Synthetic.

OS

OS Mycobacterium tuberculosis.

XX

PN W09942076-A2.

XX

PD 26-AUG-1999.

XX

XX 17-FEB-1999; 99WO-0503268.

XX

PR 05-MAY-1998; 98US-0072967.

PR

PR 18-FEB-1998; 98US-0025197.

XX

XX

PA (CORI-) CORIXA CORP.

XX

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX

XX WPI; 1999-527409/44.  
XX N-PSDB; AAZ19368.

XX

XX

PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
PT skin tests and protective or therapeutic vaccines or compositions

XX

XX Disclosure; Page 205-208; 299pp; English.

XX

CC The present invention describes polypeptides comprising an immunogenic  
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
CC other polypeptides fragments, can be used in pharmaceutical compositions  
CC or vaccines to generate a protective or therapeutic immune response to  
CC M. tuberculosis and as reagents in skin tests for diagnosis of  
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
CC by, T, B or natural killer cells and/or macrophages in  
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to  
CC AAY39225 are used in the exemplification of the present invention.

XX

SQ Sequence 802 AA;

Query Match 97.8%; Score 481; DB 20; Length 802;

Best Local Similarity 99.0%; Pred. No. 8.1e-44;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWRCAGTAQAQAAVVRFOEA 61  
Db 425 aemktdaatlaqagnferisgdlktqldqvstagslqgqrgaagtaagaavvrifgea 484  
QY 62 ANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQMGF 100  
Db 485 ankqkqeldeistnirqagvqysradeeqqalssqmgf 523

RESULT 13  
AAAY39081  
ID AAY39081 standard; Protein; 802 AA.  
XX AAY39081;  
XX 05-NOV-1999 (first entry)  
DT M tuberculosis fusion protein Tbf-6.  
DE Antigen; diagnosis; detection; infection; antibody; immunisation;  
KW vaccine; immunity.  
XX Mycobacterium tuberculosis.  
OS Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
PN WO9942118-A2.  
XX 26-AUG-1999.  
XX 17-FEB-1999; 99WO-US03265.  
XX 05-MAY-1998; 98US-0072596.  
PR 18-FEB-1998; 98US-0024753.  
XX (CORI-) CORIXA CORP.  
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
XX WPI; 1999-527416/44.  
DR N-PSDB; AAZ19245.

XX New polypeptide comprising antigenic portions of M. tuberculosis  
XX Example 10; Page 316-318; 323pp; English.  
XX This invention describes novel recombinant antigens and their encoding  
XX nucleic acids derived from Mycobacterium tuberculosis. The novel  
XX polypeptides are useful for detecting M. tuberculosis infection in a  
XX biological sample by detecting antibodies which bind with the  
XX polypeptides, and are useful as vaccines for immunizing against  
XX M. tuberculosis infection. The new detection methods are needed as  
XX current vaccination strategies do not provide 100% immunity.  
SQ Sequence 802 AA;

Query Match 97.8%; Score 481; DB 20; Length 802;  
Best Local Similarity 99.0%; Pred. No. 8.1e-44;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWRCAGTAQAQAAVVRFOEA 61  
Db 425 aemktdaatlaqagnferisgdlktqldqvstagslqgqrgaagtaagaavvrifgea 484  
QY 62 ANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQMGF 100  
Db 485 ankqkqeldeistnirqagvqysradeeqqalssqmgf 523

Query Match 97.8%; Score 481; DB 20; Length 802;  
Best Local Similarity 99.0%; Pred. No. 8.1e-44;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWRCAGTAQAQAAVVRFOEA 61  
Db 425 aemktdaatlaqagnferisgdlktqldqvstagslqgqrgaagtaagaavvrifgea 484  
QY 62 ANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQMGF 100  
Db 485 ankqkqeldeistnirqagvqysradeeqqalssqmgf 523

RESULT 14

AAAY39033  
ID AAY39033 standard; Protein; 802 AA.  
XX AAY39033;  
XX 05-NOV-1999 (first entry)  
DT M tuberculosis fusion protein Tbf-2.  
DE Antigen; diagnosis; detection; infection; antibody; immunisation;  
KW vaccine; immunity.  
XX Mycobacterium tuberculosis.  
OS Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
PN WO9942118-A2.  
XX 26-AUG-1999.  
XX 17-FEB-1999; 99WO-US03265.  
XX 05-MAY-1998; 98US-0072596.  
PR 18-FEB-1998; 98US-0024753.  
XX (CORI-) CORIXA CORP.  
XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
XX WPI; 1999-527416/44.  
DR N-PSDB; AAZ19156.

XX New polypeptide comprising antigenic portions of M. tuberculosis  
XX Example 10; Page 251-253; 323pp; English.  
XX This invention describes novel recombinant antigens and their encoding  
XX nucleic acids derived from Mycobacterium tuberculosis. The novel  
XX polypeptides are useful for detecting M. tuberculosis infection in a  
XX biological sample by detecting antibodies which bind with the  
XX polypeptides, and are useful as vaccines for immunizing against  
XX M. tuberculosis infection. The new detection methods are needed as  
XX current vaccination strategies do not provide 100% immunity.  
SQ Sequence 802 AA;

Query Match 97.8%; Score 481; DB 20; Length 802;  
Best Local Similarity 99.0%; Pred. No. 8.1e-44;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTDATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWRCAGTAQAQAAVVRFOEA 61  
Db 425 aemktdaatlaqagnferisgdlktqldqvstagslqgqrgaagtaagaavvrifgea 484  
QY 62 ANKQKQELDEISTNIRQAGVQYSRADDEQQQALSSQMGF 100  
Db 485 ankqkqeldeistnirqagvqysradeeqqalssqmgf 523

RESULT 15  
AAU01901  
ID AAU01901 standard; Protein; 983 AA.  
XX AAU01901;  
XX 29-AUG-2001 (first entry)  
DT M tuberculosis Tbf15 fusion protein.  
DE Tbf15; antigen; vaccine; tuberculosis; AIDS; His tag; Ra3; 38kd;  
KW 38-1; FL Tbf4; acquired immunodeficiency disease.  
XX Synthetic.

OS Mycobacterium tuberculosis.  
XX  
FH Key Location/Qualifiers  
FT Binding-site 3..8  
FT /label= Histidine\_tag  
FT /note= "Nickel chelating region used for purifying  
FT the fusion protein"  
FT Region 9..74  
FT /label= Ra3\_region  
FT Region 75..425  
FT /label= 38kD\_region  
FT Region 426..524  
FT /label= 38-1\_region  
FT Region 525..983  
FT /label= FL\_TbH4\_region  
XX  
PN WO200124820-A1.  
XX  
XX  
PD 12-APR-2001.  
XX  
XX 10-OCT-2000; 2000WO-US28095.  
XX  
XX 07-OCT-1999; 9BUS-0158338.  
PR 07-OCT-1999; 9BUS-0158425.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;  
XX  
XX WPI: 2001-290576/30.  
DR N-PSDB; AAS03795.  
XX  
XX  
PT Vaccinating against Mycobacteria infections in mammals using fusion  
PT proteins comprising combinations of heterologous antigens  
XX  
XX Claim 6; Fig 4; 168pp; English.  
XX  
CC The sequence represents Mycobacterium fusion protein antigen TbF15  
CC consisting of a His tag for purification, antigens Ra3, 38kD, 38-1  
CC and FL-TbH4 (full-length TbH4). Compositions comprising at least 2  
CC heterologous antigens, as a fusion protein, and vectors expressing the  
CC fusion proteins are used as vaccines to prophylactically immunise mammals  
CC (especially humans) against infection by Mycobacteria. The compositions  
CC contain at least 2 heterologous antigens that increase the serological  
CC sensitivity of individuals infected with tuberculosis, a disease  
CC frequently affecting patients with acquired immunodeficiency disease,  
CC AIDS.  
XX  
XX Sequence 983 AA;  
SQ  
  
Query Match 97.8%; Score 481; DB 22; Length 983;  
Best Local Similarity 99.0%; Pred. No. 1e-43;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 2 AEMKTDATLQDEAGNPFISGRDGLKTDQVVESTAGSLQGMGAAGTAAQAAYVRFQEA 61  
Dy 426 aemktdaatlaqeaqnferisgdkltdqdvvestagslqgmrgaagtaaqavvrifqea 485  
Qy 62 ANKOKQELDEISTNRAGVQYSRADERQOQALSSQMGF 100  
Dy 486 ankqkqeldeistnirgagvqysradeeqqalssqmgf 524

Search completed: July 22, 2002, 01:11:37  
Job time: 13424 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2002, 21:50:13 ; Search time 24.18 seconds  
(without alignments)  
101.016 Million cell updates/sec

Title: US-09-462-480-5  
Perfect score: 492  
Sequence: 1 MAEMKTDATLQGEAGNFER.....VOYSRADEEQOALSSQMGF 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	486	98.8	100	4	US-08-818-112-115
2	486	98.8	100	4	US-08-818-111-110
3	486	98.8	100	4	US-09-056-556-115
4	481	97.8	802	4	US-09-056-556-214
5	462	93.9	95	4	US-08-818-112-88
6	462	93.9	95	4	US-08-818-111-89
7	462	93.9	95	4	US-09-056-556-88
8	392	79.7	80	4	US-08-818-112-117
9	392	79.7	80	4	US-08-818-111-112
10	392	79.7	80	4	US-09-056-556-117
11	122	24.8	28	4	US-08-818-112-98
12	122	24.8	28	4	US-09-056-556-98
13	121	24.6	27	4	US-08-818-112-96
14	121	24.6	27	4	US-09-056-556-96
15	118	24.0	27	4	US-08-818-112-95
16	118	24.0	27	4	US-08-818-112-97
17	118	24.0	27	4	US-09-056-556-95
18	118	24.0	27	4	US-09-056-556-97
19	114	23.2	28	4	US-08-818-112-93
20	114	23.2	28	4	US-09-056-556-93
21	85	17.3	16	4	US-08-818-112-94
22	85	17.3	16	4	US-09-056-556-94
23	81.5	16.6	955	1	US-08-006-6768-1
24	81.5	16.6	955	1	US-08-282-845-2
25	81.5	16.6	955	2	US-08-428-414A-3
26	81.5	16.6	955	5	PCT-US94-00324-1
27	73	14.8	1147	1	US-08-144-121-3

28	73	14.8	1147	2	US-08-735-893-3	Sequence 3, Appli
29	73	14.8	1165	1	US-08-144-121-2	Sequence 2, Appli
30	73	14.8	1165	1	US-08-735-893-2	Sequence 2, Appli
31	73	14.8	2101	1	US-08-466-390-4	Sequence 4, Appli
32	73	14.8	2101	1	US-08-470-950-4	Sequence 4, Appli
33	73	14.8	2101	1	US-08-467-781-4	Sequence 4, Appli
34	73	14.8	2101	1	US-08-195-487-4	Sequence 4, Appli
35	73	14.8	2101	2	US-08-483-924-4	Sequence 4, Appli
36	73	14.8	2101	4	US-09-452-294-1	Sequence 1, Appli
37	73	14.8	2101	5	PCT-US93-06160-4	Sequence 4, Appli
38	68.5	13.9	349	6	5510466-2	Patent No. 5510466
39	68.5	13.9	453	6	5510466-4	Patent No. 5510466
40	67.5	13.7	795	1	US-07-716-827C-5	Sequence 5, Appli
41	66.5	13.5	2005	3	US-08-836-325-7	Sequence 7, Appli
42	66	13.4	84	1	US-08-452-592B-9	Sequence 9, Appli
43	66	13.4	551	2	US-09-067-351-2	Sequence 2, Appli
44	66	13.4	551	4	US-09-360-490-2	Sequence 2, Appli
45	65	13.2	84	1	US-08-452-592B-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1  
US-08-818-112-115  
; Sequence 115, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,112  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; US-08-818-112-115

Query Match 98.8%; Score 486; DB 4; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-48;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGTAQAAAVVRFQE 60  
|||||  
Db 1 MAEMKTDATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGTAQAAAVVRFQE 60  
QY 61 AANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSQMGF 100  
|||||  
Db 61 AANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSQMGF 100

## RESULT 2

US-08-818-111-110  
; Sequence 110, Application US/08818111  
; Patent No. 6338852  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,111  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 110:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-818-111-110

Query Match 98.8%; Score 486; DB 4; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-48;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGTAQAAAVVRFQE 60  
|||||  
Db 1 MAEMKTDATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGTAQAAAVVRFQE 60  
QY 61 AANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSQMGF 100  
|||||  
Db 61 AANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSQMGF 100

## RESULT 3

US-09-056-556-115  
; Sequence 115, Application US/09056556  
; Patent No. 6350456

; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,556  
; FILING DATE: 07-APR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.457  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 115:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 100 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-09-056-556-115

Query Match 98.8%; Score 486; DB 4; Length 100;  
Best Local Similarity 99.0%; Pred. No. 1.7e-48;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDATLGQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGTAQAAAVVRFQE 60  
|||||  
Db 1 MAEMKTDATLAQAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAAGTAQAAAVVRFQE 60  
QY 61 AANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSQMGF 100  
|||||  
Db 61 AANKQKQELDEISTNIRQAGVOYSRADDEEQOQALSSQMGF 100

## RESULT 4

US-09-056-556-214  
; Sequence 214, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/056,556  
;; FILING DATE: 07-APR-1998  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Maki, David J.  
;; REGISTRATION NUMBER: 31,392  
;; REFERENCE/DOCKET NUMBER: 210121.457  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (206) 622-4900  
;; TELEFAX: (206) 682-6031  
;; INFORMATION FOR SEQ ID NO: 214:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 802 amino acids  
;; TYPE: amino acids  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
US-09-056-556-214

Query Match 97.8%; Score 481; DB 4; Length 802;  
Best Local Similarity 99.0%; Pred. No. 9.7e-47;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AEMKTDATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEA 61  
Db 425 AEMKTDATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEA 484  
QY 62 ANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100  
Db 485 ANKQKQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 523

RESULT 5  
US-08-818-112-88  
; Sequence 88, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 88:  
; SEQUENCE CHARACTERISTICS:

;; LENGTH: 95 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
US-08-818-112-88

Query Match 93.9%; Score 462; DB 4; Length 95;  
Best Local Similarity 98.9%; Pred. No. 8.8e-46;  
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 6 TDAATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65  
Db 1 TDAATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60  
QY 66 KQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 100  
Db 61 KQELDEISTNIRQAGVQYSRADEEQOQALSSQMGF 95

RESULT 6  
US-08-818-111-89  
; Sequence 89, Application US/08818111  
; Patent No. 6338852  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 89:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-818-111-89

Query Match 93.9%; Score 462; DB 4; Length 95;  
Best Local Similarity 98.9%; Pred. No. 8.8e-46;  
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 6 TDAATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 65  
Db 1 TDAATLQGEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAQAAVVRFOEAANKQ 60

QY 66 KOELDEISTNIRQAGVOYSDRADEEQOQALSSQMGF 100  
|||||  
Db 61 KOELDEISTNIRQAGVOYSDRADEEQOQALSSQMGF 95  
RESULT 7  
US-09-056-556-88  
; Sequence 88, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; TREATMENT OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,556  
; FILING DATE: 07-APR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.457  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 88:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 95 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-056-556-88

Query Match 93.9%; Score 462; DB 4; Length 95;  
Best Local Similarity 98.9%; Pred. No. 8.8e-46;  
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 6 TDAATLQGEAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEAANKQ 65  
|||||  
Db 1 TDAATLQGEAGNFRISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEAANKQ 60  
QY 66 KOELDEISTNIRQAGVOYSDRADEEQOQALSSQMGF 100  
|||||  
Db 61 KOELDEISTNIRQAGVOYSDRADEEQOQALSSQMGF 95  
RESULT 8  
US-08-818-112-117  
; Sequence 117, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 153  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/818,112  
; FILING DATE: 13-MAR-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 117:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 80 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-818-112-117

Query Match 79.7%; Score 392; DB 4; Length 80;  
Best Local Similarity 100.0%; Pred. No. 6.8e-38;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 ISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEAANKQKQELDEISTNIRQAG 80  
|||||  
Db 1 ISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAVVRFOEAANKQKQELDEISTNIRQAG 60  
QY 81 VOYSDRADEEQOQALSSQMGF 100  
|||||  
Db 61 VOYSDRADEEQOQALSSQMGF 80

RESULT 9  
US-08-818-111-112  
; Sequence 112, Application US/08818111  
; Patent No. 6338852  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TUBERCULOSIS  
; NUMBER OF SEQUENCES: 148  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS



SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,111  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C6  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 112:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-111-112

Query Match 79.7%; Score 392; DB 4; Length 80;  
Best Local Similarity 100.0%; Pred. No. 6.8e-38;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 21 ISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFEAANKQKQELDEISTNIRQAG 80  
|||||  
Db 1 ISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFEAANKQKQELDEISTNIRQAG 60  
QY 81 VOYSDRADEEQQALSSQMGF 100  
|||||  
Db 61 VOYSDRADEEQQALSSQMGF 80

## RESULT 10

US-09-056-556-117  
Sequence 117, Application US/09056556  
Patent No. 6350456  
GENERAL INFORMATION:  
APPLICANT: Read, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, David C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
NUMBER OF SEQUENCES: 241  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 117:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

## US-09-056-556-117

Query Match 79.7%; Score 392; DB 4; Length 80;  
Best Local Similarity 100.0%; Pred. No. 6.8e-38;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 21 ISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFEAANKQKQELDEISTNIRQAG 80  
|||||  
Db 1 ISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFEAANKQKQELDEISTNIRQAG 60  
QY 81 VOYSDRADEEQQALSSQMGF 100  
|||||  
Db 61 VOYSDRADEEQQALSSQMGF 80

## RESULT 11

US-08-818-112-98  
Sequence 98, Application US/08818112  
Patent No. 6290969  
GENERAL INFORMATION:  
APPLICANT: Read, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, David C.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Houghton, Raymond  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,112  
FILING DATE: 13-MAR-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C6  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 98:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-818-112-98

Query Match 24.8%; Score 122; DB 4; Length 28;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 IRQAGVOYSDRADEEQQALSSQMGF 100  
|||||  
Db 4 IRQAGVOYSDRADEEQQALSSQMGF 28

## RESULT 12

TOPOLOGY: linear

ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
City: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:

US-09-056-556-96

```

; STRANDED: single
; TOPOLOGY: linear

```

**TRE**

Best Local Similarity 92.6%; Pred. NO. 1.4e-07;  
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1 GCGGTAQA AAVRFQEAANKQKQELDE 27

Matches 24; Conservative 0; Mismatches 0; Indels 0

Db 4 STAGSLQGWRGAAGTAAQAAVVR 27

Job time: 12129 sec

**THIS PAGE BLANK (uspto)**



; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/10/140,045  
; PRIOR FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: US/09/116,492A  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 100  
; TYPE: LPT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-140-045-5

Query Match 100.0%; Score 492; DB 6; Length 100;  
Best Local Similarity 100.0%; Pred. No. 7.3e-44;  
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
|||||  
Db 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
QY 61 AANKQKQELDEISTNIRAGVQVSRADDEQQQALSSQMGF 100  
|||||  
Db 61 AANKQKQELDEISTNIRAGVQVSRADDEQQQALSSQMGF 100

RESULT 3  
US-10-080-170B-639  
; Sequence 639, Application US/10080170B  
; GENERAL INFORMATION:  
; APPLICANT: COLE, P.S.T.  
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
; FILE REFERENCE: 03495.0218  
; CURRENT APPLICATION NUMBER: US/10/080,170B  
; PRIOR FILING DATE: 2002-06-10  
; PRIOR APPLICATION NUMBER: 60/270,123  
; PRIOR FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 652  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 639  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-080-170B-639

Query Match 98.8%; Score 486; DB 6; Length 100;  
Best Local Similarity 99.0%; Pred. No. 3e-43;  
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
|||||  
Db 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGTAAQAAYVRFQE 60  
QY 61 AANKQKQELDEISTNIRAGVQVSRADDEQQQALSSQMGF 100  
|||||  
Db 61 AANKQKQELDEISTNIRAGVQVSRADDEQQQALSSQMGF 100

RESULT 4  
US-09-116-492A-6  
; Sequence 6, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOB

; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492A  
; CURRENT FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 49  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492A-6

Query Match 50.2%; Score 247; DB 5; Length 49;  
Best Local Similarity 100.0%; Pred. No. 6.5e-19;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGT 49  
|||||  
Db 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGT 49  
RESULT 5  
US-10-140-045-6  
; Sequence 6, Application US/10140045  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MY  
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/10/140,045  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: US/09/116,492A  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 49  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-140-045-6

Query Match 50.2%; Score 247; DB 6; Length 49;  
Best Local Similarity 100.0%; Pred. No. 6.5e-19;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGT 49  
|||||  
Db 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGGAAGT 49  
RESULT 6  
US-09-116-492A-7  
; Sequence 7, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MY  
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, A  
; FILE REFERENCE: 0660-0137-27X

; CURRENT APPLICATION NUMBER: US/09/116,492A  
; CURRENT FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 42  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492A-7

Query Match 41.7%; Score 205; DB 5; Length 42;  
Best Local Similarity 100.0%; Pred. No. 1.2e-14;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 QEAANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100  
|||||  
DB 1 QEAANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 42

RESULT 7  
US-10-140-045-7  
; Sequence 7, Application US/10140045  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOB  
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS W  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/10/140,045  
; PRIOR FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: US/09/116,492A  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 42  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-140-045-7

Query Match 41.7%; Score 205; DB 6; Length 42;  
Best Local Similarity 100.0%; Pred. No. 1.2e-14;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 QEAANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100  
|||||  
DB 1 QEAANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 42

RESULT 8  
US-09-116-492A-28  
; Sequence 28, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOB  
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS W  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492A  
; CURRENT FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631

; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 28  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium leprae  
US-09-116-492A-28

Query Match 38.0%; Score 187; DB 5; Length 100;  
Best Local Similarity 40.0%; Pred. No. 2.6e-12;  
Matches 40; Conservative 20; Mismatches 40; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOE 60  
|||||  
DB 1 MAEMITEAAILTQAAQFDQIASGLSQERNFVDSIQSFQNTWEGQAASAAALGALGRFE 60

QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100  
|||||  
DB 61 AMQDQIRQLESIVDKLNRSGGNTTKTDEANQLLSKMN 100

RESULT 9  
US-10-140-045-28  
; Sequence 28, Application US/10140045  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MY  
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, I  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/10/140,045  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: US/09/116,492A  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 28  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium leprae  
US-10-140-045-28

Query Match 38.0%; Score 187; DB 6; Length 100;  
Best Local Similarity 40.0%; Pred. No. 2.6e-12;  
Matches 40; Conservative 20; Mismatches 40; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOE 60  
|||||  
DB 1 MAEMITEAAILTQAAQFDQIASGLSQERNFVDSIQSFQNTWEGQAASAAALGALGRFE 60

QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQALSSQMGF 100  
|||||  
DB 61 AMQDQIRQLESIVDKLNRSGGNTTKTDEANQLLSKMN 100

RESULT 10  
US-10-080-170B-12  
; Sequence 12, Application US/10080170B  
; GENERAL INFORMATION:  
; APPLICANT: COLE, S.T.  
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES  
; FILE REFERENCE: 03495.0218  
; CURRENT APPLICATION NUMBER: US/10/080,170B

; CURRENT FILING DATE: 2002-06-10  
; PRIOR APPLICATION NUMBER: 60/270,123  
; PRIOR FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 652  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Mycobacterium leprae  
US-10-080-170B-12

Query Match 38.0%; Score 187; DB 6; Length 100;  
Best Local Similarity 40.0%; Pred. No. 2.6e-12;  
Matches 40; Conservative 20; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEMKTDAAATGCEAGNFERISGLTKTQIDOVSTAGSLQGWKRGAGTAQAQAAVRFQE 60  
Db 1 MAEMITEAAILTQAAAFQOIAAGLSOERNFVDSIGSFQNTWEGQAASALGALGRFDE 60  
QY 61 AANKQKQELDEISTNIRQAGVQYSRADDEEQQAALSSQMGE 100  
Db 61 AMQDQIRQLESIVDKLRSGGNYTKTDDRANQLLSKKNF 100

RESULT 11  
US-09-116-492A-12  
; Sequence 12, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492A  
; CURRENT FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492A-12

Query Match 27.8%; Score 137; DB 5; Length 28;  
Best Local Similarity 100.0%; Pred. No. 7.6e-08;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 73 STNIRQAGVQYSRADDEEQQAALSSQMGE 100  
Db 1 STNIRQAGVQYSRADDEEQQAALSSQMGE 28

RESULT 12  
US-10-140-045-12  
; Sequence 12, Application US/10140045  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/10/140,045  
; CURRENT FILING DATE: 2002-05-08

; PRIOR APPLICATION NUMBER: US/09/116,492A  
; PRIOR FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 28  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-140-045-12

Query Match 27.8%; Score 137; DB 6; Length 28;  
Best Local Similarity 100.0%; Pred. No. 7.6e-08;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 73 STNIRQAGVQYSRADDEEQQAALSSQMGE 100  
Db 1 STNIRQAGVQYSRADDEEQQAALSSQMGE 28

RESULT 13  
US-09-116-492A-8  
; Sequence 8, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492A  
; CURRENT FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492A-8

Query Match 20.3%; Score 100; DB 5; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.00033;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 13 QEAGNFERISGDLKTKTQIDQV 32  
Db 1 QEAGNFERISGDLKTKTQIDQV 20

RESULT 14  
US-10-140-045-8  
; Sequence 8, Application US/10140045  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; TITLE OF INVENTION: USING THE SAME  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/10/140,045  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: US/09/116,492A  
; PRIOR FILING DATE: 1998-07-16  
; CURRENT APPLICATION NUMBER: 60/252,631



; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-140-045-8

Query Match 20.3%; Score 100; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.00033;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 QEAGNFERISGLKTDQDV 32  
|||||  
DB 1 QEAGNFERISGLKTDQDV 20

RESULT 15  
US-09-116-492A-10  
; Sequence 10, Application US/09116492A  
; GENERAL INFORMATION:  
; APPLICANT: GICQUEL, BRIGITTE  
; APPLICANT: BERTHET, FRANCOIS-XAVIER  
; APPLICANT: ANDERSEN, PETER  
; APPLICANT: RASMUSSEN, PETER B  
; TITLE OF INVENTION: POLYNUCLEOTIDE FUNCTIONALLY CODING FOR THE LHP PROTEIN FROM MYCOBACTERIUM TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; TITLE OF INVENTION: TUBERCULOSIS, ITS BIOLOGICALLY ACTIVE DERIVATIVE FRAGMENTS, AS WELL AS  
; FILE REFERENCE: 0660-0137-27X  
; CURRENT APPLICATION NUMBER: US/09/116,492A  
; CURRENT FILING DATE: 1998-07-16  
; PRIOR APPLICATION NUMBER: 60/252,631  
; PRIOR FILING DATE: 1997-07-16  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-116-492A-10

Query Match 16.1%; Score 79; DB 5; Length 16;  
Best Local Similarity 100.0%; Pred. No. 0.037;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 GSLQGQWRGAGTAA 51  
|||||  
DB 1 GSLQGQWRGAGTAA 15

Search completed: July 22, 2002, 01:18:01  
Job time: 382 sec

THIS PAGE BLANK (USPTO)



```
Query Match      38.0%; Score 187; DB 2; Length 100;
Best Local Similarity 40.0%; Pred. No. 3.7e-10;
Matches 40; Conservative 20; Mismatches 40; Indels 0; Gaps 0;

QY 1 MAEMKTDATLTGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOE 60
Db 1 MAEMITEAAILTQOAAQFOIAGSLGSOERNFVDSIQSFQNTWEGQAASAAALGALGRFDE 60

QY 61|AANKQKQELDEISTNIRQAGVOYSRADDEQQOALSOMGF 100
Db 61 AMQDQIRQLESIVDKLNRSNGVTKTDDANQLLSKMMNF 100

RESULT 3
T35661
probable chromosome associated protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 02-Jun-2000
C:Accession: T35661
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z21585
A:Accession: T35661
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-1186 <MUR>
A:Cross-references: EMBL:AL034447; PIDN:CAA22420.1; GSPDB:GN00070; SCOEDB:SC7A1.21
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC7A1.21
C:Superfamily: chromosome segregation protein SMC1

Query Match      18.8%; Score 92.5; DB 2; Length 1186;
Best Local Similarity 32.3%; Pred. No. 1.8;
Matches 32; Conservative 16; Mismatches 44; Indels 7; Gaps 3;

QY 2 AEMKTDATLTGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTA--AQAAVVRFO 59
Db 686 AGRRRECAALVELGERRAADREKSSVAQ---QLGLAGQARGAGAEERSAAARAQ 742

QY 60 EAANKQKQELDEISTNIRQAGVOYSRADDEQQOALSQM 98
Db 743 EALDKALTEVEELAE--RLVAEEMPEVEEDPTAARDRL 779

RESULT 4
B59103
hypothetical protein pX01-98 - Bacillus anthracis virulence plasmid pX01
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: B59103
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler
J.; Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harbored
A:Reference number: A59091; MUID:99445483
A:Accession: B59103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32402.1; PID:g4894314
A:Experimental source: strain Sterne
C:Genetics:
A:Gene: pX01-98
A:Genome: plasmid

Query Match      18.6%; Score 91.5; DB 2; Length 410;
Best Local Similarity 28.0%; Pred. No. 0.68;
Matches 26; Conservative 17; Mismatches 29; Indels 21; Gaps 3;

QY 3 EMKTDATLTGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAA 62
```

```
Db 11 QLEQAARTVKNTRSSLEYTHODLYSQTEYIAS-----QMSGASSD-----REYQMF 56
QY 63 NKQK-----QELDEISTNIRQAGVOYSRADDE 88
Db 57 NEAKPMFNIQLQELDKIAVELERAARVFRADDE 89

RESULT 5
T38435
coiled coil protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38435
R:McDougall, R.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z21793
A:Accession: T38435
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-750 <MCD>
A:Cross-references: EMBL:AL009227; PIDN:CAA15821.1; GSPDB:GN00066; SPDB:SPAC27D7.02c
A:Experimental source: strain 972h-; cosmid c27D7
C:Genetics:
A:Gene: SPDB:SPAC27D7.02c
A:Map position: 1

Query Match      18.3%; Score 90; DB 2; Length 750;
Best Local Similarity 28.2%; Pred. No. 1.8;
Matches 24; Conservative 18; Mismatches 23; Indels 20; Gaps 3;

QY 13 QEAG--NFERISGLTKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAAANKQKQELD 70
Db 550 KOAGENHYNLSDDYETQIKSLESSL-----TNSQAEVSVFQEKINELNSQID 597

QY 71 EISTNIRQAGVOYSRADDEQQOALS 95
Db 598 ELKLNLEANKKY-----QELAIS 616

RESULT 6
B70600
hypothetical protein Rv3905c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70600
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: B70600
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-103 <COL>
A:Cross-references: GB:Z94121; GB:AL123456; NID:g3261736; PIDN:CAB08096.1; PID:e31227
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv3905c

Query Match      18.1%; Score 89; DB 2; Length 103;
Best Local Similarity 23.1%; Pred. No. 0.25;
Matches 21; Conservative 22; Mismatches 48; Indels 0; Gaps 0;

QY 4 MKTDAATLTGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOEAA 63
Db 7 LRVEPAVMQGFASLDGAAEHLAVQLAELDAQVQGLGMRGAGSGAYSAGWELWHRGAG 66

QY 64 KQKQELDEISTNIRQAGVOYSRADDEQQOAL 94
```

[illegible]



Search completed: July 22, 2002, 01:13:26  
Job time: 10123 sec

THIS PAGE BLANK (USPTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 01:13:29 ; Search time 23.86 seconds  
(without alignments)  
162.278 Million cell updates/sec

Title: US-09-462-480-5  
Perfect score: 492  
Sequence: 1 MAEMKTAATLQEGAGNER.....VOYRADEQQQALSSQMGF 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	481	97.8	99	1	CF10_MYCTU
2	182	37.0	99	1	CF10_MYCLE
3	89	18.1	103	1	ES6D_MYCTU
4	87	17.7	2022	1	ANT1_ONCVO
5	81.5	16.6	955	1	KINL_LEICH
6	79	16.1	245	1	YT27_MYCLE
7	78.5	16.0	1960	1	MYH9_HUMAN
8	76	15.4	491	1	K2M2_SHEEP
9	75.5	15.3	440	1	FENR_ANAVA
10	75.5	15.3	1961	1	MYH9_RAT
11	75	15.2	2349	1	TPR_HUMAN
12	74.5	15.1	80	1	PYS1_ANASP
13	74.5	15.1	242	1	GRPE_HALME
14	74	15.0	1325	1	G160_MOUSE
15	73.5	14.9	440	1	FENR_ANASO
16	73.5	14.9	440	1	FENR_ANASP
17	73.5	14.9	975	1	KINH_DROME
18	73	14.8	1589	1	PHP_DROME
19	72.5	14.7	80	1	PYS1_MASLA
20	72.5	14.7	575	1	FLA2_CAMJE
21	72.5	14.7	1238	1	SBCC_RHOCA
22	72.5	14.7	1679	1	YIO9_YEAST
23	72	14.6	373	1	CL1D_HUMAN
24	72	14.6	502	1	K2M3_SHEEP
25	71.5	14.5	292	1	BPHC_PSES1
26	71.5	14.5	1959	1	MYH9_CHICK
27	71	14.4	457	1	MESE_LEUME
28	71	14.4	548	1	HLXB_VIBCH
29	70.5	14.3	78	1	PYS1_SYNEL
30	70.5	14.3	845	1	Y4FA_RHISN
31	70	14.2	284	1	TPM2_SCHMA
32	70	14.2	512	1	MCPD_ENTAE
33	70	14.2	535	1	HTR1_HALN1

34	70	14.2	576	1	PEX5_PICPA
35	70	14.2	715	1	CLPB_MYCPN
36	69.5	14.1	1172	1	LMB3_HUMAN
37	69	14.0	186	1	APL3_GALME
38	69	14.0	189	1	APL3_MANSE
39	69	14.0	1938	1	MYSD_CAEEL
40	68.5	13.9	356	1	IRPA_SYNP7
41	68.5	13.9	453	1	MSRE_BOVIN
42	68.5	13.9	456	1	RMUC_XYLFA
43	68.5	13.9	1107	1	VJEP_ECOLI
44	68.5	13.9	1742	1	MY5C_HUMAN
45	68.5	13.9	1947	1	MY5C_CAEEL

## ALIGNMENTS

RESULT 1

ID	CF10_MYCTU	STANDARD;	PRT;	99 AA.
AC	O69739;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	10 kDa culture filtrate antigen cfp10.			
GN	CFP10 OR LHP OR RV3874 OR MT3988 OR MT027.09.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98061212; PubMed=9846755;			
RA	Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,			
RA	Gicquel B.;			
RT	"A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel			
RL	low-molecular-mass culture filtrate protein (CFP-10).";			
RL	Microbiology 144:3195-3203(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Hoiroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence.";			
RL	Nature 393:537-544(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bisai W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains.;"			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
CC	-!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			

CC or send an email to license@isb-sib.ch).

DR EMBL; AF004671; AAC83445.1; -;  
DR EMBL; AL022120; CAA17966.1; -;  
DR EMBL; AE007190; AAK48356.1; -;

DR TIGR; MT3988; -;

DR TuberculList; RV3874; -;

KW Complete proteome.

FT INIT\_MET 0

SQ SEQUENCE 99 AA; 10663 MW; EBCEAE6A996C5489D CRC64;

#### Query Match

Best Local Similarity 97.8%; Score 481; DB 1; Length 99;

Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTDAAATLGOEAGNFERISGDLTKTQIDOVSTAGSLQGWGGAAGTAAQAAYVRFQEA 61

Db 1 AEMKTDAAATLGOEAGNFERISGDLTKTQIDOVSTAGSLQGWGGAAGTAAQAAYVRFQEA 60

QY 62 ANKQKQELDEISTNRQAGVQYSRADEEQOQALSSOMGF 100

Db 61 ANKQKQELDEISTNRQAGVQYSRADEEQOQALSSOMGF 99

#### RESULT 2

CF10\_MYCLE

ID CF10\_MYCLE STANDARD; PRT; 99 AA.

AC O33084;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE 10 kDa culture filtrate antigen cfp10 homolog.

GN ML0050 OR MLCB628.13C.

OS Mycobacterium leprae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

ON NCBI\_TaxID=1769;

RP SEQUENCE FROM N.A.

RA Eiglmeyer K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;

RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

RT [2]

RN SEQUENCE FROM N.A.

RC STRAIN=TN;

RX MEDLINE=21128732; PubMed=11234002;

RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,

RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,

RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,

RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,

RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,

RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,

RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,

RA Barrell B.G.;

RT "Massive gene decay in the leprosy bacillus.";

RL Nature 409:1007-1011(2001).

CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Y14967; CAA75210.1; -;

DR EMBL; AL583917; CAC29558.1; -;

DR Leproma; ML0050; -;

KW Complete proteome.

FT INIT\_MET 0

SQ SEQUENCE 99 AA; 10833 MW; 859B484F7EF5A8A CRC64;

BY SIMILARITY

Query Match 37.0%; Score 182; DB 1; Length 99;

Best Local Similarity 39.4%; Pred. No. 3.9e-10;

Matches 39; Conservative 20; Mismatches 40; Indels 0; Gaps 0;

QY 2 AEMKTDAAATLGOEAGNFERISGDLTKTQIDOVSTAGSLQGWGGAAGTAAQAAYVRFQEA 61

Db 1 AEMITEAAILTQQAQFDQISGLSQERNFVDSIGSFQNTWEGQAASALGALGRFDEA 60

QY 62 ANKQKQELDEISTNRQAGVQYSRADEEQOQALSSOMGF 100

Db 61 MODQIRQLESIYDKLNRSNGNYTKTDDANQLLSKMNMF 99

#### RESULT 3

ES6D\_MYCTU

ID ES6D\_MYCTU STANDARD; PRT; 103 AA.

AC O05440;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative ESAT-6 like protein 13.

GN RV3905C OR MT4024 OR MTCY15F10.06.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

ON NCBI\_TaxID=1773;

RP SEQUENCE FROM N.A.

RN STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence.";

RL Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains";

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO THE ESAT6 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Z94121; CAB08096.1; -;

DR EMBL; AE007193; AAK48388.1; -;

DR TIGR; MT4024; -;

DR TuberculList; RV3905C; -;

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 103 AA; 10460 MW; 3994E272A7BDF02 CRC64;

Query Match

18.1%; Score 89; DB 1; Length 103;

[illegible]

## RESULTS

RESULT 6

MEDLINE=20057165; PubMed=10591208;  
Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,  
Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,  
Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,  
Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,  
Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,  
Coble V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,  
Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C.,  
Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,  
Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,  
Graham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,  
Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,  
Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,  
Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,  
Matthews L., McCann O.T., Mcclay J., McLaren S., McMurray A.A.,  
Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,  
Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,  
Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,  
Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,  
Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,  
Whiteley M.N., Willey D., Williams L., Williams S., Williamson D.R.,  
Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,  
Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,  
Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,  
Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,  
Do A., Do T., Dorman A., Fang F., Fu Y., Hua A., Kenton S.,  
Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaï E.,  
Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,  
Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.  
Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,  
Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,  
Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,  
Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,  
Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,  
Wilson R., Korf I., Bedel J.A., Hillier L., Hardis E., Waterston K.,  
Nelson J., Emanuel B.S., Shaikh T., Kurahashi H., Saïtta S.,  
Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,  
Payard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
Khan A.S., Lane L., Tilahun Y., Wright H.;  
The DNA sequence of human chromosome 22.;  
Nature 402:489-495(1999).  
[2]  
SEQUENCE OF 1-1337 FROM N.A.  
MEDLINE=92003925; PubMed=1912569;  
Toothaker L.E., Gonzalez D.A., Tung N., Lemons R.S., le Beau M.M.,  
Araout M.A., Clayton L.K., Tenen D.G.;  
"Cellular myosin heavy chain in human leukocytes: isolation of 5'  
cdna clones, characterization of the protein, chromosomal  
localization, and upregulation during myeloid differentiation.";  
Blood 78:1826-1833(1991).  
[3]  
SEQUENCE OF 1-715 FROM N.A.  
MEDLINE=913136803; PubMed=1960190;  
Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,  
Gdula D., Adelstein R.S., Weir L.;  
"Human nonmuscle myosin heavy chains are encoded by two genes located  
on different chromosomes.";  
Circ. Res. 69:530-539(1991).  
[4]  
SEQUENCE OF 714-1960 FROM N.A.  
MEDLINE=90138958; PubMed=1967836;  
Saez C.G., Myers J.C., Shows T.B., Leinwand L.A.;  
"Human nonmuscle myosin heavy chain mRNA: generation of diversity  
through alternative polyadenylation.";  
proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).  
[5]  
VARIANT DFNA17 HIS-705.  
MEDLINE=20489856; PubMed=11023810;  
Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,  
Mhatre A.N.;  
"Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in

RT nonmuscle myosin MYH9.";

RL Am. J. Hum. Genet. 67:1121-1128(2000).

RN [6]

RP VARIANTS MHA/FTNS/SBS K-93; C-702; C-1165; H-1424 AND K-1841.

RX MEDLINE-20428192; PubMed-10973259;

RA Seri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,

RA Chiglieri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apollito M.,

RA Iolascon A., Zelanite L.L., Savio A., Balduini C.L., Norris P.,

RA Magrini U., Belletti S., Heath K.E., Babcock M., Glucksman M.J.,

RA Aliprandis E., Bizzaro N., Desnick R.J., Martignetti J.A.;

RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and

RT Sebastian syndromes.";

RL Nat. Genet. 26:103-105(2000).

RN [7]

RP VARIANTS MHA ILE-1155 AND LYS-1841.

RX MEDLINE-20428193; PubMed-10973260;

RA Kelley M.J., Jawlin W., Ortel T.L., Korczak J.F.;

RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in

RT May-Hegglin anomaly.";

RL Nat. Genet. 26:106-108(2000).

CC - FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,

CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND

CC CAPPING.

CC - SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY

CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2

CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC - DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC - DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY

CC (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED

CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.

CC - DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF FECHTNER SYNDROME

CC (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED

CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.

CC WITH ADDITIONAL ALPORT-LIKE CLINICAL FEATURES OF SENSORINEURAL

CC DEAFNESS, CATARACTS AND NEPHRITIS.

CC - DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME

CC (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED

CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUKOCYTE INCLUSIONS.

CC - DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT

CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA17) WHICH IS

CC CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND

CC COCHLEOSACULAR DEGENERATION.

CC - SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC - SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC This [SWISS-PROT] entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; 282215; CAB05105.1; -

DR EMBL; M81105; AAA59888.1; -

DR EMBL; M69180; AAA61765.1; -

DR EMBL; M31013; AAA36349.1; -

DR HSP; P08799; ILV. -

DR MIM; 160775; -

DR MIM; 153640; -

DR MIM; 155100; -

DR MIM; 603622; -

DR MIM; 603249; -

DR InterPro; IPR000048; IQ.

DR InterPro; IPR004009; Myosin\_N.

DR InterPro; IPR002928; Myosin\_tail.

DR InterPro; IPR002017; Spectrin.

DR InterPro; IPR001609; myosin\_head.

DR Pfam; PF00612; IQ; 1.

DR Pfam; PF00063; myosin\_head; 1.

DR Pfam; PF02736; Myosin\_N; 1.

DR pfam: PF01576; Myosin\_tail; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR PRODom; PD000355; myosin\_head; 1.

DR SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.

KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;

KW Coiled coil; Alkylation; Multigene family; Disease mutation;

KW Deafness.

FT DOMAIN 1 778 MYOSIN HEAD-LIKE.

FT DOMAIN 779 808 IQ.

FT DOMAIN 837 1926 COILED COIL (POTENTIAL).

FT NP\_BIND 174 181 ATP (POTENTIAL).

FT DOMAIN 654 676 ACTIN-BINDING.

FT MOD\_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).

FT MOD\_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).

FT VARIANT 93 93 N -> K (IN MHA).

FT VARIANT 702 702 /FTID-VAR\_010791.

FT VARIANT 705 705 R -> C (IN FTNS).

FT VARIANT 1155 1155 /FTID-VAR\_010792.

FT VARIANT 1165 1165 R -> H (IN DFNA17).

FT VARIANT 1424 1424 T -> I (IN MHA).

FT VARIANT 1841 1841 /FTID-VAR\_010793.

FT CONFLICT 53 55 EAI -> RGH (IN REF. 3).

FT CONFLICT 660 660 T -> S (IN REF. 3).

FT CONFLICT 869 869 T -> M (IN REF. 4).

FT CONFLICT 931 931 C -> Y (IN REF. 4).

FT CONFLICT 1240 1241 KG -> GR (IN REF. 4).

FT CONFLICT 1350 1350 E -> EE (IN REF. 2).

FT CONFLICT 1764 1764 T -> A (IN REF. 2).

FT CONFLICT 1771 1771 S -> G (IN REF. 2).

FT SEQUENCE 1960 AA; 226531 MW; 588F84BB8C106E6F CRC64;

Query Match 16.0%; Score 78.5; DB 1; Length 1960;

Best Local Similarity 21.7%; Pred. No. 19;

Matches 34; Conservative 23; Mismatches 29; Indels 71; Gaps 5;

QY 11 LGQEGNFRISGDLK---TQIDQV-----EST 35

Db 1738 LEEEGNTELDNRKANKQIDQVTDNLERSHAQKNENARQQLERQNKELVKLQEM 1797

QY 36 AGSLQGWGGAAGTAAQAAYVRFQ-----AANKQ----- 65

Db 1798 EGVKSKYK-ASITALEAKIALEEQLDNETKERQAACKQVRRTEKKLKDVLQVDDERR 1856

QY 66 -----KQELDEISTNTRQAGVYSRADEEQQAALSSQ 97

Db 1857 NAEQYKDAQDKASTRLKQLKROLEAEAEAEAAQRANASR 1893

RESULT 8

K2M2\_SHEEP

ID K2M2\_SHEEP STANDARD; PRT; 491 AA.

AC P15241;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Keratin, type II microfibrillar, component 7C.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Caprinae; Ovis.

OX NCBI\_TaxID=9940;

RN [1]

RP SEQUENCE.

RX MEDLINE=90026244; PubMed=2803231;  
RA Sparrow L.G., Robinson C.P., McMahon D.T.W., Rubira M.R.;  
RT "The amino acid sequence of component 7c, a type II intermediate-  
filament protein from wool.";  
RL Blochem. J. 261:1015-1022(1989).  
CC -!- FUNCTION: WOOL MICROFIBRILLAR KERATIN  
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,  
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS  
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.  
CC -!- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE  
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO  
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I  
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,  
CC 7A, 7B, AND 7C).  
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
DR PIR; S05408; S05408.  
DR InterPro: IPR001664; IF.  
DR InterPro: IPR003054; Keratin\_II.  
DR Pfam: PF00038; filament; 1.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Keratin.  
FT MOD\_RES 1 1 BLOCKED.  
FT DOMAIN 1 109 HEAD.  
FT DOMAIN 110 416 ROD.  
FT DOMAIN 417 491 TAIL.  
FT DOMAIN 110 144 COIL 1A.  
FT DOMAIN 145 154 LINKER 1.  
FT DOMAIN 155 255 COIL 1B.  
FT DOMAIN 256 272 LINKER 12.  
FT DOMAIN 273 416 COIL 2.  
FT VARIANT 74 74 C -> G OR S.  
FT VARIANT 80 80 C -> S.  
FT VARIANT 144 144 F -> Y.  
FT VARIANT 232 232 S -> V.  
FT VARIANT 276 276 C -> D OR N.  
FT VARIANT 284 284 Q -> H.  
FT UNSURE 1 2 CG -> GC.  
SQ SEQUENCE 491 AA; 53681 MW; A801771FE3831ABE CRC64;

Query Match 15.4%; Score 76; DB 1; Length 491;  
Best Local Similarity 22.9%; Pred. No. 7.4;  
Matches 19; Conservative 23; Mismatches 39; Indels 2; Gaps 1;  
QY 17 NFERISGLKLTQDVSTAGSLQGWGAGTAQAQAAVVFQEAANKQKQELDEISTNI 76  
Db 273 NMDCIVAEIKAQYDDIASRAEASWYSKCEIKATVIRHGETLRRTKKEINELNRVI 332

QY 77 RQ-ACVQYSRADEEQQALSSQ 97  
Db 333 QRLTAEVENAKONSKLEAAVTQ 355

RESULT 9  
FENR\_ANAVA STANDARD; PRT; 440 AA.  
AC Q44549;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Ferredoxin--NADP reductase (EC 1.18.1.2) (FNR).  
GN PETH.  
OS Anabaena variabilis.  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
OX NCBI\_TaxID=1172;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PCC 7937 / ATCC 29413;  
RA Mannan R.M., Matthijs H.C.P., Pakrasi H.B.;  
RT "Cloning and molecular characterization of the peth gene in the  
cyanobacterium Anabaena variabilis ATCC 29413.";

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized  
CC ferredoxin + NADPH.  
CC -!- COFACTOR: FAD.  
CC -!- SUBCELLULAR LOCATION: MAY BE BOUND TO THE THYLAKOID MEMBRANE OR  
CC ANCHORED TO THE THYLAKOID-BOUND PHYCABILISOMES.  
CC -!- SIMILARITY: IN THE N-TERMINAL REGION, TO PHYCOBILISOME LINKER  
CC PROTEIN CPD.  
CC -!- SIMILARITY: WITH OTHER SPECIES FNR.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L26346; AAA91046.1; .  
DR HSSP; P21890; IQUE.  
DR InterPro: IPR001685; CpcD.  
DR InterPro: IPR003097; FAD-binding.  
DR InterPro: IPR001709; Flavopyrid\_cyt\_redctse.  
DR InterPro: IPR001433; Oxidored\_FAD.  
DR Pfam; PF01383; CpcD; 1.  
DR Pfam; PF00667; FAD-binding; 1.  
DR Pfam; PF00175; NAD-binding; 1.  
DR PRINTS; PR00371; FPNCR.  
DR PRODOM; PD002828; CpcD; 1.  
KW Oxidoreductase; Flavoprotein; NADP; FAD; Thylakoid; Membrane;  
KW Phycobilisome.  
FT DOMAIN 1 97 CPCD-LIKE.  
FT NP\_BIND 288 306 NADP (RIBOSE PART) (BY SIMILARITY).  
SQ SEQUENCE 440 AA; 48826 MW; 660BAA2DCF59BBB6 CRC64;

Query Match 15.3%; Score 75.5; DB 1; Length 440;  
Best Local Similarity 30.0%; Pred. No. 7.3;  
Matches 18; Conservative 15; Mismatches 22; Indels 5; Gaps 1;  
QY 38 SLQGWGAGTAQAQAAVVFQEAANKQKQELDEISTNI 92  
Db 2 SNOGAPEGAANVESGSRVFEVVGVRQNEETDQTNPIRKSGVFIRPYNRMNOMOR 61

RESULT 10  
MYH9\_RAT STANDARD; PRT; 1961 AA.  
ID MYH9\_RAT  
AC Q62812;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,  
DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).  
GN MYH9.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
CC CAPPING.  
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2  
CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2002, 01:12:24 ; Search time 66.82 seconds  
(without alignments)

258.897 Million cell updates/sec

Title: US-09-462-480-5

Perfect score: 492

Sequence: 1 MAEMKTDATLQGEAGNFER.....VOYRADEPQQALSSQMGF 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phage.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	92.5	18.8	1186	2 Q9ZBQ2	Q9ZBQ2 streptomyc
2	91.5	18.6	410	2 Q9X367	Q9X367 bacillus an
3	90	18.3	750	3 O42657	O42657 schizosacch
4	87.5	17.8	97	16 O53692	O53692 mycobacteri
5	84.5	17.2	97	16 O53267	O53267 mycobacteri
6	83.5	17.0	97	16 Q92FQ4	Q92FQ4 listeria in
7	83	16.9	1200	11 Q921B9	Q921B9 mus musculu
8	82.5	16.8	188	5 O77248	O77248 spodoptera
9	81	16.5	528	5 Q26589	Q26589 schistosoma
10	81	16.5	1940	5 Q02456	Q02456 schistosoma
11	79.5	16.2	163	16 Q9RY46	Q9RY46 deinozococcus
12	79.5	16.2	1999	11 Q63731	Q63731 rattus norv
13	79	16.1	508	6 Q28582	Q28582 ovis aries
14	78	15.9	103	16 Q97M22	Q97M22 clostridium
15	77	15.7	178	5 O02415	O02415 agrius conv
16	77	15.7	507	4 P78386	P78386 homo sapien

17	77	15.7	507	4	Q9NSB1	Q9nsb1 homo sapien
18	77	15.7	561	16	Q910I6	Q910i6 pseudomonas
19	77	15.7	1326	2	Q9L2C3	Q9l2c3 streptomyce
20	77	15.7	3908	5	Q9BK91	Q9bk91 strongyloce
21	76.5	15.5	528	16	Q9RSJ1	Q9rsj1 deinococcus
22	76.5	15.5	536	5	Q95R41	Q95r41 drosophila
23	76.5	15.5	707	2	Q9F5N1	Q9f5n1 rhizobium m
24	76.5	15.5	842	5	Q9VEH0	Q9veh0 drosophila
25	76.5	15.5	874	5	Q960Y8	Q960y8 drosophila
26	76.5	15.5	1278	9	Q9XJA8	Q9xja8 streptococc
27	76	15.4	331	2	Q9X7H6	Q9x7h6 paracoccus
28	76	15.4	1530	4	O43241	O43241 homo sapien
29	76	15.4	2756	10	Q9LJ60	Q9lj60 arabidopsi
30	75.5	15.3	529	11	Q99J33	Q99jj3 mus musculu
31	75.5	15.3	6713	16	Q99U54	Q99u54 staphylococ
32	75.5	15.3	6713	16	Q931R6	Q931r6 staphylococ
33	75	15.2	507	11	Q9D7M4	Q9d7m4 mus musculu
34	75	15.2	507	11	Q922T6	Q922t6 mus musculu
35	75	15.2	671	13	Q9YHD4	Q9yhd4 rana catesb
36	74.5	15.1	387	2	Q54842	Q54842 streptococc
37	74.5	15.1	548	2	Q9EWY9	Q9ewy9 streptomyce
38	74.5	15.1	576	2	Q9R950	Q9r950 campylobact
39	74.5	15.1	789	2	Q93M73	Q93m73 xanthomonas
40	74.5	15.1	880	16	Q9I245	Q9i245 pseudomonas
41	74.5	15.1	929	3	P78718	P78718 haematonect
42	74.5	15.1	1313	10	Q9XIP6	Q9xip6 arabidopsi
43	74.5	15.1	1617	5	Q95YX4	Q95yx4 leishmania
44	74.5	15.1	1956	5	O20641	O20641 caenorhabdi
45	74	15.0	493	4	Q9NSB3	Q9nsb3 homo sapien

## ALIGNMENTS

RESULT 1

ID	Q9ZBQ2	PRELIMINARY;	PRT; 1186 AA.
AC	Q9ZBQ2;		
DT	01-MAY-1999 (TReMBLrel. 10, Created)		
DT	01-MAY-1999 (TReMBLrel. 10, Last sequence update)		
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)		
DE	PUTATIVE CHROMOSOME ASSOCIATED PROTEIN.		
GN	SC7A1.21.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycetaceae; Streptomycetes.		
OX	NCBI_TaxID=1902;		
[1]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
RA	Murphy L., Harris D.;		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
[2]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
RA	Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;		
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.		
[3]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN=A3(2);		
RA	Redenbach M., Klesner H.M., Denapaita D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;		
RT	"A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";		
RL	Mol. Microbiol. 21:77-96(1996).		
DR	EMBL; AL034447; CAA22420.1; -		
DR	InterPro; IPR003439; ABC_transportr.		
DR	InterPro; IPR001687; ATP_Gtp_A.		
DR	InterPro; IPR003405; SMC_C.		
DR	InterPro; IPR003395; SMC_N.		
DR	InterPro; IPR003662; sub_transporter.		
DR	Pfam; PF02483; SMC_C; 1.		



RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.	
RA	Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,	
RA	Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,	
RA	Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,	
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,	
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,	
RT	"Comparative genomics of <i>Listeria species</i> ."	
RT	Science 294:849-852(2001).	
DR	EMBL; AL596163; CAC95282.1; -	
DR	LR Listlist; LIN00049; -	
KW	Complete proteome.	
SQ	SEQUENCE 97 AA; 11342 MW; 5CEB89D26C1BA775 CRC64;	
	Query Match 17.0%; Score 83.5; DB 16; Length 97;	
	Best Local Similarity 25.5%; Pred. NO. 2.3;	
	Matches 26; Conservative 20; Mismatches 39; Indels 17; Gaps	
QY	2 AEMKTDATLQGEAGNFERISGDLKTDIOVESTAGSLOGWRGAAGTAQAAVVRQEA 61	
DB	10 SELRDRAKTYGSGRDIE---DILSLSLQD---QLRSEWEG-----QAFMRFPDQ 55	
QY	62 ANKQKQELDEISTNIRAGVOYSR---ADEEQOQALSSQMFG 100	
DB	56 FQQLAPKPYTEFANLMDQINDOLEKTANAVEHQDQLSQNGFG 97	
RESULT	7	
Q921B9	PRELIMINARY;	PRT; 1200 AA.
ID	Q921B9	
AC	Q921B9;	
DT	01-DEC-2001 (TrEMBLrel. 19, Created)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE	NUCLEAR PORE COMPLEX-ASSOCIATED PROTEIN TPR (FRAGMENT).	
GN	TPR.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
OX	NCBI_TaxID=10090;	
ON	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=129/Ola;	
RC	Sandblad L., Hunziker A., Cordes V.C.;	
RT	"Evolutionarily conserved mouse tpr is a single-copy gene located on	
RT	chromosome 1.;"	
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AJ298076; CAC40701.1; -	
FT	NON_TER 1200 1200	
SQ	SEQUENCE 1200 AA; 138973 MW; 616663FBA6B267D7 CRC64;	
	Query Match 16.9%; Score 83; DB 11; Length 1200;	
	Best Local Similarity 27.18; Pred. No. 41;	
	Matches 29; Conservative 16; Mismatches 38; Indels 24; Gaps	
QY	7 DAATLQGEAGNFE-----RISG-----DLKTQIDQVESTAGSLOGWRGAA 47	
DB	900 DIATLKQHLNNNEAQLASQSTQRTGKGPGDRDDVDLKSQLRQAEEQVNDLKERLKTST 95	
QY	48 GTAQ--AAVRFQPAANKQKQELDEISTNIRAGVOYSRADEEQQ 92	
DB	960 SNVEQYRAMVTSLDSLNKKEQVTEEVHKNIET---VRLKESAETFQ 1003	
RESULT	8	
O77248	PRELIMINARY;	PRT; 188 AA.
ID	O77248	
AC	O77248;	
DT	01-NOV-1998 (TrEMBLrel. 08, Created)	
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE	APOLIPOPHORIN-III.	



```
AC Q9RV46;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE HYPOTHETICAL 17.8 KDA PROTEIN.
GN DR0105.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vanathavan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001873; AAF09701.1; -.
DR TIGR; DR0105; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 163 AA; 17830 MW; A0AC12CE20534D9C CRC64;

Query Match 16.2%; Score 79.5; DB 16; Length 163;
Best Local Similarity 22.7%; Pred. No. 9.3;
Matches 22; Conservative 16; Mismatches 46; Indels 13; Gaps 1;

QY 11 LQGEAGN-----FERISGLKTDQIDVESTAGSLQGWGGAAGTAAQAAVVR 57
Db 1 WDEAGNAQDRAEAAAARAARASHGADYREMDSNDDLLSLGLGQAKHKTDQAAEKVRAD 60

QY 58 FOEAANKQKQELDEISTNIRQAGVYSRADEEQQAL 94
Db 61 TQEAQAQNAKQADVVRANVHESAQDFRAGAQAQTL 97

RESULT 12
Q63731
ID Q63731 PRELIMINARY; PRT; 1999 AA.
AC Q63731;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE NEURONAL MYOSIN HEAVY CHAIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=92235856; PubMed=1569576;
RA Sun W., Chantler P.D.;
RT "Cloning of the cDNA encoding a neuronal myosin heavy chain from
RT mammalian brain and its differential expression within the central
RT nervous system[see comments].";
RL J. Mol. Biol. 224:1185-1193(1992).
DR HSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 2.
```

```
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin.
SQ SEQUENCE 1999 AA; 231456 MW; 090C181D55727B41 CRC64;

Query Match 16.2%; Score 79.5; DB 11; Length 1999;
Best Local Similarity 23.1%; Pred. No. 1.5e+02;
Matches 31; Conservative 26; Mismatches 28; Indels 49; Gaps 5;

QY 9 ATLIGQEAGN-----FERISGLKTDQIDVESTAGSLQGWGGAAGTAAQAAVVR 58
Db 1764 ADNLNLRGHAKQNAARQQLERQNKELVKVQEME---GTVKSKYK-ASITALEAKIAQL 1819

QY 59 QE-----AANKQ-----KQELDEISTNIRQAGVQY 83
Db 1820 EQQLDNETKERQAACKQVRTEKKLKDVLQVDERRNAEQYKDQADKASTRLKQLKROL 1879

QY 84 SRADPEEQQALSSO 97
Db 1880 EEAEAEARQANASR 1893

RESULT 13
Q28582
ID Q28582 PRELIMINARY; PRT; 508 AA.
AC Q28582;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HAIR TYPE II KERATIN INTERMEDIATE FILAMENT PROTEIN.
GN KII-9.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92274852; PubMed=1375545;
RA Powell B., Crocker L., Rogers G.;
RT "Hair follicle differentiation: expression, structure and
RT evolutionary conservation of the hair type II keratin intermediate
RT filament gene family.";
RL Development 114:417-433(1992).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR EMBL; X62509; CAA44368.1; -.
DR InterPro; IPR001664; IF.
DR InterPro; IPR003054; Keratin_II.
DR Pfam; PF00038; filament; 1.
DR PRINTS; PR01276; TYPE2KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament; Keratin.
SQ SEQUENCE 508 AA; 55245 MW; 3E728AA9512F96F7 CRC64;

Query Match 16.1%; Score 79; DB 6; Length 508;
Best Local Similarity 22.9%; Pred. No. 36;
Matches 19; Conservative 23; Mismatches 39; Indels 2; Gaps 1;

QY 17 NFERISGLKTDQIDVESTAGSLQGWGGAAGTAAQAAVVRFOEAANKQKQELDEISTNI 76
Db 271 NMDNIVAEIKAQYDDIASRAEASWYSKCEEIKATVIRHGETLRRRTKEENELNRVI 330

QY 77 RO--AGVQYSRADEEQQALSSQ 97
Db 331 QRLTAEVENAKCONSKLEAAVTO 353

RESULT 14
```

Q97M22 PRELIMINARY; PRT; 103 AA.  
AC Q97M22;  
DT 01-OCT-2001 (TREMBlrel. 18, Created)  
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
DE UNCHARACTERIZED SMALL CONSERVED PROTEIN, HOMOLOG OF YFJA/YUKE  
DE B.SUBTILIS.  
GN CAC0047.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum";  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL; AE007518; AAK78034.1; -;  
KW Complete proteome.  
SQ SEQUENCE 103 AA; 11764 MW; 41B2E22D3DF36D6D CRC64;

Query Match 15.9%; Score 78; DB 16; Length 103;  
Best Local Similarity 21.3%; Pred. No. 7.6;  
Matches 19; Conservative 24; Mismatches 44; Indels 2; Gaps 2;

QY 3 EMKDAATLGQAGNFERISGDLKTDIDOVSTAGSLQG-QWRGAAGTA-AQAAVVRFOE 60  
DB 6 DIKINWETLQSAINEYKKNCKTTLNLLKESGLKALETEWKGAKAEAFVNAQPPNFK 65  
QY 61 AANKQKQELDEISTNIRQAGVYRADEE 89  
DB 66 GMORHCNMIGELIKELQETAKETKLDNE 94

RESULT 15  
002415 PRELIMINARY; PRT; 178 AA.  
AC 002415;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE APOLIPOPHORIN-III (FRAGMENT).  
OS Agrius convolvuli (morning glory sphinx moth).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Sphingioidea; Sphingidae; Sphinginae; Agrius.  
OX NCBI\_TaxID=55055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ragland J., Yamauchi Y., Sato R., Wells M.A., Hamano K., Tsuchide K.;  
RT "cDNA and Deduced Amino Acid Sequence of Apolipophorin-III from Agrius  
convolvuli".  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF001632; AAB61280.1; -;  
FT CHAIN 13 178 APOLIPOPHORIN-III.  
NON\_TER 1  
SQ SEQUENCE 178 AA; 19648 MW; E4DA3C4DA4341E85 CRC64;

Query Match 15.7%; Score 77; DB 5; Length 178;  
Best Local Similarity 25.0%; Pred. No. 17;  
Matches 21; Conservative 21; Mismatches 42; Indels 0; Gaps 0;

QY 11 LGQAGNFERISGDLKTDIDOVSTAGSLQGQWRGAAGTAQAAVVRFOEAANKQKQELD 70  
DB 11 LGQAGNFERISGDLKTDIDOVSTAGSLQGQWRGAAGTAQAAVVRFOEAANKQKQELD 70

Search completed: July 22, 2002, 01:19:31  
Job time: 427 sec

Db 88 LEQARQNVKTAELRKAHPEVEKEANALKDKLQAAVQTTVQESQKLAKEVASNMETNQ 147  
QY 71 EISTNIRQAGVYRADEEQQOAL 94  
DB 148 KLAPKIKQAYDDFVKQAEVQKKL 171



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:41:11 ; Search time 130.35 Seconds  
(without alignments)  
46.509 Million cell updates/sec

Title: US-09-462-480-5  
Perfect score: 492  
Sequence: 1 MAEMTDATLQEAQNGFER.....VQYSRADBEQQALSSQMGF 100

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues  
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601:\*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:\*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	492	100.0	100	20	AAV03705
2	486	98.8	100	19	AAW81706
3	486	98.8	100	19	AAW64339
4	486	98.8	100	20	AAV39136
5	486	98.8	100	20	AAV38993
6	486	98.8	100	22	AAW35218
7	486	98.8	100	22	AAW19845
8	481	97.8	802	19	AAW81746
9	481	97.8	802	19	AAW64379
10	481	97.8	802	20	AAV39063
11	481	97.8	802	20	AAV39224

12	481	97.8	802	20	AAV39176	M. tuberculosis fu
13	481	97.8	802	20	AAV39081	M. tuberculosis fu
14	481	97.8	802	20	AAV39033	M. tuberculosis fu
15	462	93.9	95	18	AAW32444	Mycobacterium tube
16	462	93.9	95	19	AAW32376	Mycobacterium tube
17	462	93.9	95	19	AAW81747	M. tuberculosis im
18	462	93.9	95	19	AAW64321	Mycobacterium tube
19	462	93.9	95	20	AAV32097	Mycobacterium tube
20	462	93.9	95	20	AAV39118	M. tuberculosis an
21	462	93.9	95	20	AAV38981	M. tuberculosis re
22	392	79.7	80	18	AAW32454	Mycobacterium tube
23	392	79.7	80	18	AAW32386	Mycobacterium tube
24	392	79.7	80	19	AAW81707	M. tuberculosis im
25	392	79.7	80	19	AAW64340	Mycobacterium tube
26	392	79.7	80	20	AAV39137	Mycobacterium tube
27	392	79.7	80	20	AAV38994	M. tuberculosis an
28	247	50.2	49	20	AAV03706	M. tuberculosis re
29	205	41.7	42	20	AAV03707	M. tuberculosis re
30	137	27.8	28	20	AAV03712	M. tuberculosis re
31	122	24.8	28	18	AAW32460	Mycobacterium tube
32	122	24.8	28	19	AAW81698	M. tuberculosis im
33	122	24.8	28	20	AAV39128	Mycobacterium tube
34	121	24.6	27	18	AAW32457	M. tuberculosis an
35	121	24.6	27	19	AAW81696	Mycobacterium tube
36	121	24.6	27	20	AAV39126	M. tuberculosis im
37	118	24.0	27	18	AAW32457	Mycobacterium tube
38	118	24.0	27	18	AAW32459	Mycobacterium tube
39	118	24.0	27	19	AAW81695	Mycobacterium tube
40	118	24.0	27	19	AAW81697	M. tuberculosis im
41	118	24.0	27	20	AAV39125	M. tuberculosis im
42	118	24.0	27	20	AAV39127	M. tuberculosis an
43	114	23.2	28	18	AAW32455	Mycobacterium tube
44	114	23.2	28	19	AAW81693	M. tuberculosis im
45	114	23.2	28	20	AAV39123	M. tuberculosis an

## ALIGNMENTS

RESULT 1  
AAV03705  
ID AAY03705 standard; Protein; 100 AA.  
XX  
AC AAY03705;  
XX  
DT 07-JUN-1999 (first entry)  
XX  
DE M. tuberculosis LHP polypeptide.  
XX  
KW ESAT-6 protein; antigenic protein; LHP; Mycobacterium tuberculosis;  
KW immunogen; vaccine; bacteria; tuberculosis; regulatory; synergistic;  
KW immune response.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN WO9904005-A1.  
XX  
PD 28-JAN-1999.  
XX  
PF 16-JUL-1998; 98WO-IB01091.  
XX  
PR 16-JUL-1997; 97US-0052631.  
XX  
PA (INSP ) INST PASTEUR.  
XX (STAT-) STATENS SERUM INST.  
PI Andersen P, Berthet F, Gicquel B, Rasmussen PB;  
XX WPI; 1999-132249/11.  
DR N-PSDB; AAX29168, AAX29171.  
XX  
PT New nucleic acid containing regulator and LHP gene of Mycobacterium tuberculosis - useful in vaccines, for diagnosis, and for expression

PT of heterologous proteins

XX Claim 17; Page 64; 88pp; English.

XX The present invention is directed to a polynucleotide carrying the  
 CC regulatory expression signals of the ESAT-6 protein as well as an open  
 CC reading frame coding for an antigenic protein LHP from Mycobacterium  
 CC tuberculosis. Host cells comprising the polynucleotide are used for the  
 CC recombinant expression of the protein. The recombinant polypeptide can be  
 CC used as immunogens and vaccines, to protect against bacteria of the  
 CC M. tuberculosis complex in humans or animals (the vaccines may include  
 CC other immunogenic proteins of the bacteria or their fragments,  
 CC specifically ESAT-6); and (b) for diagnosing tuberculosis infection by  
 CC detection of specific antibodies. The regulatory region present in the  
 CC polynucleotide may be used to express almost any heterologous protein in  
 CC mycobacteria, particularly as a fusion with polyhistidine. The two  
 CC proteins encoded by the polynucleotide, LHP and ESAT-6, are expected to  
 CC provide a synergistic increase in ability to induce a protective immune  
 CC response. The present sequence represents the LHP polypeptide.

XX Sequence 100 AA;

Query Match 100.0%; Score 492; DB 20; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-46;  
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWRCAGTAQAQAAVRFQE 60  
 Db 1 maemktdaaatlgaggnferisgdlktqldqvvestagslqgwgagtaagaavvrifqe 60  
 QY 61 AANKQKQELDEISTNRQAGVQYSRADDEQQOALSSQMGF 100  
 Db 61 aankkqkqeldeistnrqagvgysradeeqqalssqmgf 100

RESULT 2

AAW81706  
 ID AAW81706 standard; Protein; 100 AA.

XX AAW81706;

XX 27-JAN-1999 (first entry)

XX M. tuberculosis immunogenic polypeptide Tb38-IN.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 XX vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX WO9816646-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18293.

XX 13-MAR-1997; 97US-0818112.

XX 11-OCT-1996; 96US-0730510.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-261042/23.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX Example 3B; Page 138-139; 230pp; English.

CC This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This  
 CC sequence can be formulated into vaccines and/or pharmaceutical  
 CC compositions for immunising against M. tuberculosis infection or may  
 CC be used for the diagnosis of tuberculosis.

XX Sequence 100 AA;

Query Match 98.8%; Score 486; DB 19; Length 100;  
 Best Local Similarity 99.0%; Pred. No. 9.6e-46;  
 Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWRCAGTAQAQAAVRFQE 60  
 Db 1 maemktdaaatlgaggnferisgdlktqldqvvestagslqgwgagtaagaavvrifqe 60  
 QY 61 AANKQKQELDEISTNRQAGVQYSRADDEQQOALSSQMGF 100  
 Db 61 aankkqkqeldeistnrqagvgysradeeqqalssqmgf 100

RESULT 3

AAW64339  
 ID AAW64339 standard; Protein; 100 AA.

XX AAW64339;

XX 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen Tb38-IN.

XX Tuberculosis; infection; diagnosis; antigen; Tb37-FL.

XX Mycobacterium tuberculosis strain H37Rv.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

XX 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and diagnosis of tuberculosis

XX Example 3; Page 145; 250pp; English.

XX This polypeptide comprises a partial sequence of Mycobacterium  
 CC tuberculosis antigen Tb38-IN. It is encoded by genomic DNA isolated  
 CC from a M. tuberculosis strain H37Rv genomic library using a probe  
 CC derived from clone Tb38-1 (see AAW44384). The invention relates to  
 CC compositions and methods for diagnosing tuberculosis. It provides  
 CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of an  
 CC a soluble M. tuberculosis antigen, or an immunogenic portion of an  
 CC M. tuberculosis antigen, as well as DNA sequences encoding such  
 CC polypeptides, recombinant expression vectors and transformed or  
 CC transfected host cells. Also claimed are methods and diagnostic  
 CC kits for detecting M. tuberculosis infection in a patient using  
 CC these polypeptides, antibodies or oligonucleotide probes and  
 CC primers, for the diagnosis of tuberculosis.

SQ Sequence 100 AA;

Query Match 98.8%; Score 486; DB 19; Length 100;  
 Best Local Similarity 99.0%; Pred. No. 9.6e-46;  
 Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWRGAAGTAAGAAVVRFE 60  
 DB 1 maemktdaatlaqeagngferisgdlktqldqvvestagslqgwrngaagtaagaavvrfe 60

QY 61 AANKQKQELDEISTNIRAGVQVYSRADEEQQALSSQMGE 100  
 DB 61 aankkqkqeldeistniragvgvysradeeqqalssqmgf 100

RESULT 4  
 AAY39136  
 ID AAY39136 standard; Protein; 100 AA.  
 XX  
 AC AAY39136;  
 DT 05-NOV-1999 (first entry)  
 DE M. tuberculosis antigen Tb38-IN amino acid sequence.  
 XX  
 KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;  
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;  
 KW immune response; skin test.  
 XX  
 OS Mycobacterium tuberculosis.  
 PN WO9942076-A2.  
 PD 26-AUG-1999.  
 PF 17-FEB-1999; 99WO-US03268.  
 PR 05-MAY-1998; 98US-0072967.  
 PR 18-FEB-1998; 98US-0025197.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1999-527409/44.  
 XX  
 PT New antigens from Mycobacterium tuberculosis useful in diagnostic  
 PT skin tests and protective or therapeutic vaccines or compositions  
 PS Example 3; Page 133-134; 299pp; English.  
 XX  
 CC The present invention describes polypeptides comprising an immunogenic  
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described  
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.  
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and  
 CC other polypeptides fragments, can be used in pharmaceutical compositions  
 CC or vaccines to generate a protective or therapeutic immune response to  
 CC M. tuberculosis and as reagents in skin tests for diagnosis of  
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion  
 CC by, T, B or natural killer cells and/or macrophages in  
 CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to  
 CC AAY39225 are used in the exemplification of the present invention.  
 XX  
 SQ Sequence 100 AA;

Query Match 98.8%; Score 486; DB 20; Length 100;  
 Best Local Similarity 99.0%; Pred. No. 9.6e-46;  
 Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWRGAAGTAAGAAVVRFE 60

DB 1 maemktdaatlaqeagngferisgdlktqldqvvestagslqgwrngaagtaagaavvrfe 60

QY 61 AANKQKQELDEISTNIRAGVQVYSRADEEQQALSSQMGE 100  
 DB 61 aankkqkqeldeistniragvgvysradeeqqalssqmgf 100

RESULT 5  
 AAY38993  
 ID AAY38993 standard; Protein; 100 AA.  
 XX  
 AC AAY38993;  
 DT 05-NOV-1999 (first entry)  
 DE M. tuberculosis recombinant antigen protein Tb38-IN.  
 XX  
 KW Antigen; diagnosis; detection; infection; antibody; immunisation;  
 KW vaccine; immunity.  
 XX  
 OS Mycobacterium tuberculosis.  
 PN WO9942118-A2.  
 PD 26-AUG-1999.  
 PF 17-FEB-1999; 99WO-US03265.  
 PR 05-MAY-1998; 98US-0072596.  
 PR 18-FEB-1998; 98US-0024753.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;  
 PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX  
 DR WPI; 1999-527416/44.  
 XX  
 PT New polypeptide comprising antigenic portions of M. tuberculosis  
 PS Example 3; Page 179; 323pp; English.  
 XX  
 CC This invention describes novel recombinant antigens and their encoding  
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel  
 CC polypeptides are useful for detecting M. tuberculosis infection in a  
 CC biological sample by detecting antibodies which bind with the  
 CC polypeptides, and are useful as vaccines for immunizing against  
 CC M. tuberculosis infection. The new detection methods are needed as  
 CC current vaccination strategies do not provide 100% immunity.  
 XX  
 SQ Sequence 100 AA;

Query Match 98.8%; Score 486; DB 20; Length 100;  
 Best Local Similarity 99.0%; Pred. No. 9.6e-46;  
 Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWRGAAGTAAGAAVVRFE 60  
 DB 1 maemktdaatlaqeagngferisgdlktqldqvvestagslqgwrngaagtaagaavvrfe 60

QY 61 AANKQKQELDEISTNIRAGVQVYSRADEEQQALSSQMGE 100  
 DB 61 aankkqkqeldeistniragvgvysradeeqqalssqmgf 100

RESULT 6  
 AAB35218  
 ID AAB35218 standard; Protein; 100 AA.  
 XX  
 AC AAB35218;  
 XX

DT 24-APR-2001 (first entry)  
XX M tuberculosis Rv3874 protein.  
XX  
XX  
XX Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c;  
KW Rv1037c; Rv2346c; Rv2653c; Rv2654c; Rv3020c; Rv3444c;  
KW Rv3445c; Rv3890c; Rv3891c; Rv3904c; Rv3905c.  
XX  
XX Mycobacterium tuberculosis.  
OS  
XX  
XX WO200104151-A2.  
PN  
XX  
XX 18-JAN-2001.  
PD  
XX  
XX 13-JUL-2000; 2000WO-DK00398.  
PF  
XX  
XX 13-JUL-1999; 99DK-0001070.  
PR  
XX 15-JUL-1999; 99US-0144011.  
PR  
XX  
XX (STAT-) STATENS SERUM INST.  
PA  
XX  
XX Andersen P, Skjot R;  
PI  
XX  
XX WPI; 2001-091923/10.  
DR  
XX  
XX New polypeptide encoded by a member of the esat-6-gene family for  
PT immunizing against and diagnosis of tuberculosis.  
PT  
XX  
XX Example 2; Page 65; 80pp; English.  
PS  
XX  
XX The present invention provides the protein and coding sequences for  
CC members of the esat-6 gene family from Mycobacterium tuberculosis. These  
CC proteins include Rv0287, Rv1037c, Rv2346c, Rv2348c, Rv2653c,  
CC Rv2654c, Rv3020c, Rv3444c, Rv3445c, Rv3890c, Rv3891c, Rv3904c and  
CC Rv3905c. These can be used to produce vaccines against, and in the  
CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of  
CC the proteins of the invention.  
XX  
SQ Sequence 100 AA;  
  
Query Match 98.8%; Score 486; DB 22; Length 100;  
Best Local Similarity 99.0%; Pred. No. 9.6e-46;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MAEMKTDATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAAAGTAAQAAVVRFE 60  
Db 1 maemktdaatlaqeagnferisgdlktqidqvvestagslqgwgaaagtaaqaaavvrfe 60  
  
QY 61 AANKQKQELDEISTNIRAGVQYSRADDEEQQALSSQMGF 100  
Db 61 aankqkgeldeistniragvgvysradeeqqalssqmgf 100  
  
RESULT 7  
AAB19845  
ID AAB19845 standard; Protein; 100 AA.  
XX  
XX AAB19845;  
AC  
XX  
XX 05-MAR-2001 (first entry)  
DT  
XX  
XX Mycobacterium tuberculosis protein MTBN4.  
DE  
XX  
XX MTBN4; tuberculosis; BCG; vaccine; infection; diagnosis.  
KW  
XX  
XX Mycobacterium tuberculosis.  
OS  
XX  
XX WO200066157-A1.  
PN  
XX  
XX 09-NOV-2000.  
PD  
XX  
XX 04-MAY-2000; 2000WO-US12257.  
PF

XX 04-MAY-1999; 99US-0132505.  
PR  
XX  
XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.  
PA  
XX  
XX Gennaro ML;  
PI  
XX  
XX WPI; 2001-007153/01.  
DR  
XX  
XX N-PSDB; AAA89038.  
DR  
XX  
XX Novel polypeptide encoded by open reading frames present in  
PT Mycobacterium tuberculosis genome and not by the BCG strain of M.  
PT bovis, useful as vaccine and for diagnosing tuberculosis infection  
PT  
XX  
XX Claim 11; Fig 1; 35pp; English.  
PS  
XX  
XX The present sequence is that of the Mycobacterium tuberculosis  
CC MTBN4 protein. This is 1 of 8 proteins, i.e. MTBN1-8 (see  
CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)  
CC identified as being present in the genome of M. tuberculosis but  
CC absent from the genome of the BCG strain of Mycobacterium bovis.  
CC MTBN1-8 represent reagents that are useful in discriminating between  
CC M. tuberculosis and BCG and, in particular, for diagnostic methods  
CC which discriminate between exposure of a subject to M. tuberculosis  
CC and vaccination with BCG. The invention features these MTBN  
CC polypeptides, functional fragments of them, DNA encoding them,  
CC vectors, transformed cells, and diagnostic, therapeutic, and  
CC prophylactic (vaccine) methods, including genetic vaccination  
CC methods.  
XX  
SQ Sequence 100 AA;  
  
Query Match 98.8%; Score 486; DB 22; Length 100;  
Best Local Similarity 99.0%; Pred. No. 9.6e-46;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MAEMKTDATLGOEAGNFERISGLTKTQIDQVESTAGSLQGWGAAAGTAAQAAVVRFE 60  
Db 1 maemktdaatlaqeagnferisgdlktqidqvvestagslqgwgaaagtaaqaaavvrfe 60  
  
QY 61 AANKQKQELDEISTNIRAGVQYSRADDEEQQALSSQMGF 100  
Db 61 aankqkgeldeistniragvgvysradeeqqalssqmgf 100  
  
RESULT 8  
AAW81746  
ID AAW81746 standard; Protein; 802 AA.  
XX  
XX AAW81746;  
AC  
XX  
XX 27-JAN-1999 (first entry)  
DT  
XX  
XX M. tuberculosis fusion protein Tbf-2.  
DE  
XX  
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis.  
KW  
XX  
XX Synthetic.  
OS  
XX Mycobacterium tuberculosis.  
OS  
XX  
XX WO9816646-A2.  
PN  
XX  
XX 23-APR-1998.  
PD  
XX  
XX 07-OCT-1997; 97WO-US18293.  
PF  
XX  
XX 13-MAR-1997; 97US-0818112.  
XX  
XX 11-OCT-1996; 96US-0730510.  
PR  
XX  
XX (CORI-) CORIAXA CORP.  
PA  
XX

PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
DR WPI; 1998-261042/23.  
DR N-PSDB; AAW64567.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and for diagnosis, treatment and prevention of tuberculosis

XX Disclosure; Page 208-211; 230pp; English.

XX This sequence represents the fusion protein Tbf-2 which is composed of  
CC immunogenic polypeptides from Mycobacterium tuberculosis (MT). This  
CC protein is used in a method for inducing protective immunity against  
CC tuberculosis (TB). This sequence can be formulated into vaccines  
CC and/or pharmaceutical compositions for immunising against  
CC M. tuberculosis infection or may be used for the diagnosis of TB.

XX Sequence 802 AA;

Query Match 97.8%; Score 481; DB 19; Length 802;

Best Local Similarity 99.0%; Pred. No. 4.3e-44;

Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTAATLGQAGNFERISGDLKTQIDQVESTAGSLQGWRCAGTAQAQAAVVRFOEA 61

||||| 425 aemktdaatlaqaagnferisgdlktqldqvstagslqgwgaaagaaavvrfoea 484

QY 62 ANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 100

||||| 485 ankqkqeldeistnirqagvqysradeeqqalssqmgf 523

RESULT 9

AAW64379

ID AAW64379 standard; Protein; 802 AA.

XX

AC AAW64379;

XX 09-NOV-1998 (first entry)

XX Mycobacterium antigen Tbf2 protein fusion.

XX Tuberculosis; infection; diagnosis; 38 kDa antigen; TBra3; DPEP;

XX Tb38-1; Tbf-2.

XX Mycobacterium tuberculosis.

XX Synthetic.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US18214.

XX 13-MAR-1997; 97US-0818111.

XX 11-OCT-1996; 96US-0729622.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;

PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;

XX WPI; 1998-251292/22.

DR N-PSDB; AAW55801.

XX New Isolated Mycobacterium tuberculosis polypeptides and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PT and diagnosis of tuberculosis

XX Example 7; Page 223-226; 250pp; English.

XX

CC This polypeptide comprises a fusion protein, designated Tbf-2,  
CC composed of Mycobacterium tuberculosis antigens TBra3 (see AAW64295),  
CC 38 kDa antigen (see AAW64364), Tb38-1 (see AAW64321) and DPEP (see  
CC AAW64322). It was produced by PCR amplification (see AAW4450-57) of  
CC the appropriate antigen DNA sequences, cloning into an expression  
CC vector, and expression in E. coli. Tbf-2 can be used for  
CC serodiagnosis of tuberculosis. The invention relates to  
CC compositions and methods for diagnosing tuberculosis. It provides  
CC polypeptides (see AAW64291-W64379) comprising antigenic or  
CC immunogenic portions of M. tuberculosis antigens, or fusion proteins,  
CC DNA sequences encoding such polypeptides, recombinant expression  
CC vectors and host cells. Also claimed are methods and diagnostic  
CC kits for detecting M. tuberculosis infection in a patient.

XX Sequence 802 AA;

Query Match 97.8%; Score 481; DB 19; Length 802;

Best Local Similarity 99.0%; Pred. No. 4.3e-44;

Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEMKTAATLGQAGNFERISGDLKTQIDQVESTAGSLQGWRCAGTAQAQAAVVRFOEA 61

||||| 425 aemktdaatlaqaagnferisgdlktqldqvstagslqgwgaaagaaavvrfoea 484

QY 62 ANKQKQELDEISTNIRQAGVQYSRADDEEQQALSSQMGF 100

||||| 485 ankqkqeldeistnirqagvqysradeeqqalssqmgf 523

RESULT 10

AAAY32063

ID AAY32063 standard; Protein; 802 AA.

XX

AC AAY32063;

XX 17-JAN-2000 (first entry)

XX Mycobacterium tuberculosis antigen fusion protein Tbf-2.

XX Tuberculosis; antigen; fusion protein; Tbf-2; TBra3; 38kD; Tb38-1;

XX DPEP; diagnosis; therapy; vaccine; immunogen.

XX Mycobacterium tuberculosis.

XX WO9951748-A2.

XX 14-OCT-1999.

XX 07-APR-1999; 99WO-US07717.

XX 07-APR-1998; 98US-0056556.

XX 30-DEC-1998; 98US-0223040.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Alderson M, Campos-Neto A;

XX WPI; 1999-601610/51.

XX N-PSDB; AAZ20198.

XX New fusion proteins useful for diagnosis, prevention and treatment of

XX tuberculosis -

XX Claim 1; Fig 5G-J; 83pp; English.

XX This sequence represents a recombinant Mycobacterium tuberculosis  
CC tetra-antigen fusion protein, termed Tbf-2, composed of the antigens  
CC TBra3, 39kD, Tb38-1 and DPEP. The fusion protein is expressed in  
CC host cells using a vector carrying a polynucleotide (see AAZ20198)  
CC comprising the 4 coding sequences. The invention provides fusion  
CC proteins (see AAY32059-71) containing at least 2 M. tuberculosis

The present invention describes polypeptides comprising an immunogenic part of a Mycobacterium tuberculosis antigen (Ag). Also described are vaccines and fusion protein containing M. tuberculosis Ag's. M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other polypeptides fragments, can be used in pharmaceutical compositions or vaccines to generate a protective or therapeutic immune response to M. tuberculosis and as reagents in skin tests for diagnosis of tuberculosis. Ag can induce proliferation of, or cytokine secretion by, T, B or natural killer cells and/or macrophages in tuberculosis-immune subjects. AA219249 and AA219460 and AA33083 to

97.88: score 481: DB 20: Length 802:



```

XX WO9709428-A2.
XX
XX
XX PD 13-MAR-1997.
XX
XX PF 30-AUG-1996; 96WO-US14674.
XX
XX PR 12-JUL-1996; 96US-0680574.
XX PR 01-SEP-1995; 95US-0523436.
XX PR 22-SEP-1995; 95US-0533634.
XX PR 22-MAR-1996; 96US-0620874.
XX PR 05-JUN-1996; 96US-0659683.
XX
XX (CORI-) CORIXA CORP.
XX
XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
XX Twardzik DR, Vedvick TH;
XX
XX DR WPI; 1997-192903/17.
XX DR N-PSDB; AAT91509.
XX
XX PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
XX PT useful in vaccines for prevention or treatment of tuberculosis, also
XX PT for diagnosis
XX
XX PS Example 3; Page 124; 168pp; English.
XX
XX A new immunogenic polypeptide has been developed comprising an
XX immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
XX its variant differing only in conservative substitutions and/or
XX modifications). The present sequence represents a M.tuberculosis
XX antigen, Tb38-1. The immunogenic protein, and fusion proteins
XX containing one or more of the proteins or one of the proteins plus
XX ESAT-6, are useful in vaccines, preferably when formulated with a
XX non-specific adjuvant, to induce an immune response against
XX M.tuberculosis (for treatment or prevention).
XX
XX Sequence 95 AA:

```

```

Query Match 93.98; Score 462; DB 18; Length 95;
Best Local Similarity 90.98; Pred. No. 3.7e-43;
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 TDAATLQGEAGNFERISGLTKTQIDQVESTAGSLQGGOWRGAAGTAAQAAVVRFOEAANKQ 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 tdaatlacaggnferisgdlktqldqvvestagslqgqwrgaagtaaqaaavvrfoeaankq 60

Qy 66 KOELDEISTNTRQAGVQVSRADDEQQQALSSQMGF 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 kgeideistnirqagvqvsradeeqqqalssqmgf 95

```

Search completed: July 5, 2001, 11:45:42  
Job time: 271 sec





Db 912 ELASQLEATAAKSSAEQDR--ENTRAALEQQLRDSEERAAELASQ 955

## RESULT 2

US-08-282-845-2  
; Sequence 2, Application US/08282845  
; Patent No. 5719263  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: A 230kd Antigen Present in Leishmania  
; MEDIUM TYPE: Species  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple Macintosh Operating System 7.1  
; SOFTWARE: Microsoft Word for Macintosh 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/282.845  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/006,676  
; FILING DATE: JANUARY 15, 1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia Anne  
; REGISTRATION NUMBER: 34,693  
; REFERENCE/DOCKET NUMBER: 5004-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)587-0430  
; TELEFAX: (206)233-0644  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 955 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-282-845-2

Query Match 16.6%; Score 81.5; DB 1; Length 955;  
Best Local Similarity 29.2%; Pred. No. 0.4;  
Matches 31; Conservative 15; Mismatches 47; Indels 13; Gaps 2;

QY 3 EMKTDATLGOEAGNFERISGLTKTDQV-----ESTAGSLQGWGGAAGTAA 51  
Db 852 DRESTRATLEQQLRESEERAAELASQLESTTAAKMSAEQDRSTRATLEQQLRDSEERAA 911  
QY 52 QAAVVFQEAANKQKQELDEISTNIRQAGVYSRADEEQOQALSSQ 97  
Db 912 ELASQLEATAAKSSAEQDR--ENTRAALEQQLRDSEERAAELASQ 955

## RESULT 3

US-08-428-414A-3  
; Sequence 3, Application US/08428414A  
; Patent No. 5912166  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; TITLE OF INVENTION: LEISHMANIASIS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/428.414A  
; FILING DATE: 21-APR-1995  
; CLASSIFICATION: 436  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kadlecsek, Ann T.  
; REGISTRATION NUMBER: 39,244  
; REFERENCE/DOCKET NUMBER: 210121.407  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANDBERRY  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 955 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-428-414A-3

Query Match 16.6%; Score 81.5; DB 2; Length 955;  
Best Local Similarity 29.2%; Pred. No. 0.4;  
Matches 31; Conservative 15; Mismatches 47; Indels 13; Gaps 2;

QY 3 EMKTDATLGOEAGNFERISGLTKTDQV-----ESTAGSLQGWGGAAGTAA 51  
Db 852 DRESTRATLEQQLRESEERAAELASQLESTTAAKMSAEQDRSTRATLEQQLRDSEERAA 911  
QY 52 QAAVVFQEAANKQKQELDEISTNIRQAGVYSRADEEQOQALSSQ 97  
Db 912 ELASQLEATAAKSSAEQDR--ENTRAALEQQLRDSEERAAELASQ 955

## RESULT 4

PCT-US94-00324-1  
; Sequence 1, Application PC/TUS9400324  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven  
; TITLE OF INVENTION: Diagnosis of Leishmaniasis  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple System 7.1  
; SOFTWARE: Microsoft Word, version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/00324  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/006,676  
; FILING DATE: 15-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia Anne  
; REGISTRATION NUMBER: 34,693  
; REFERENCE/DOCKET NUMBER: 5004-WO

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 955 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-00324-1

Query Match 16.6%; Score 81.5; DB 5; Length 955;  
Best Local Similarity 29.2%; Pred. No. 0.4;  
Matches 31; Conservative 15; Mismatches 47; Indels 13; Gaps 2;  
PCT-US94-00324-1

QY 3 EMTDAATLGOEAGNERISGDLKTQIDQV-----ESTAGSLQGWGAAGTAA 51  
Db 852 DRESTRATLEQQLRESEERAAELASQLESTAAKMSAEQDRESTRATLEQQLRDSERAA 911  
QY 52 QAAVVRFOEAANKQKQELDEISTNIRAGVOYYSRADEEQQALSSQ 97  
Db 912 ELASQLEATRAAKSSAEQDR--ENTRAALEQQLRDSERAAELASQ 955

## RESULT 5

US-08-144-121-3  
Sequence 3, Application US/08144121  
Patent No. 5610031  
GENERAL INFORMATION:  
APPLICANT: Burgeson, Robert E.  
APPLICANT: Wagman, David W.  
TITLE OF INVENTION: BLK CHAIN OF LAMININ AND METHODS OF USE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: BOSTON  
STATE: Massachusetts  
COUNTRY: United States  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/144,121  
FILING DATE: 27-OCT-1993  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: (MGP-0780.0) MGP-021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1147 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1..231  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 232..411  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 412..765

FEATURE:  
NAME/KEY: Domain  
LOCATION: 766..1147  
US-08-144-121-3

Query Match 14.8%; Score 73; DB 1; Length 1147;  
Best Local Similarity 18.5%; Pred. No. 4.8;  
Matches 29; Conservative 23; Mismatches 45; Indels 60; Gaps 4;

QY 4 MKTDAATLGOEAGNERISG-----DLKTDQIDVESTAGSLQGW 43  
Db 898 LPTDSATVLQKMEIQAIARLPNDVLVSQTQKDIGGARRLQAEAEARSRAHVEGV 957  
QY 44 RGAAGTAAGAA-----VVRFOEAANKQK-----Q 67  
Db 958 EDVVGNLROGTVALQEAQDMQTSRLRIQDRVAEVQVLGQQLVSMTKQLGDFWT 1017  
QY 68 ELDEISTNIRAGVOYYSRADE-----EQQALSSQMGF 100  
Db 1018 RMEELRHQARQGAEAQVQAQQLAEGASEQALSQEGF 1054

## RESULT 6

US-08-735-893-3  
Sequence 3, Application US/08735893  
Patent No. 5914317  
GENERAL INFORMATION:  
APPLICANT: Burgeson, Robert E.  
APPLICANT: Wagman, David W.  
TITLE OF INVENTION: BLK CHAIN OF LAMININ AND METHODS OF USE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, suite 510  
CITY: BOSTON  
STATE: Massachusetts  
COUNTRY: United States  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/735,893  
FILING DATE: 18-OCT-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/144,121  
FILING DATE: 27-OCT-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: (MGP-0780.1) MGP-021DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1147 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1..231  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 232..411  
FEATURE:



US-08-466-390-4  
; Sequence 4, Application US/08456390  
; Patent No. 5686562  
; GENERAL INFORMATION:  
; APPLICANT: TOUTATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,390  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-466-390-4

Query Match 14.8%; Score 73; DB 1; Length 2101;  
Best Local Similarity 25.0%; Pred. No. 10;  
Matches 21; Conservative 19; Mismatches 32; Indels 12; Gaps 2;  
QY 10 TLQGEAGNFRISGDLKTQIDQVESTAGSLQGWGRGAAGTAQAQAAVYRFOEANKQKQEL 69  
|| ||| : | : : : : | : : | : | : || : || :  
Db 417 TLKQEAATLAANTQIARVEMLETERGQGEAKLLAERG-----HFEE-----EKQQL 464  
QY 70 DEISTNIRQAGVQYSRADDEQQA 93  
: | : : : | : | : || : || :  
Db 465 SSLITDLOSSISNLSQAKELEQA 488  
US-08-466-390-4  
Query Match 14.8%; Score 73; DB 1; Length 2101;  
Best Local Similarity 25.0%; Pred. No. 10;  
Matches 21; Conservative 19; Mismatches 32; Indels 12; Gaps 2;  
QY 10 TLQGEAGNFRISGDLKTQIDQVESTAGSLQGWGRGAAGTAQAQAAVYRFOEANKQKQEL 69  
|| ||| : | : : : : | : : | : | : || : || :  
Db 417 TLKQEAATLAANTQIARVEMLETERGQGEAKLLAERG-----HFEE-----EKQQL 464  
QY 70 DEISTNIRQAGVQYSRADDEQQA 93  
: | : : : | : | : || : || :  
Db 465 SSLITDLOSSISNLSQAKELEQA 488  
US-08-466-390-4  
RESULT 10  
US-08-470-950-4  
; Sequence 4, Application US/08470950  
; Patent No. 5698439  
; GENERAL INFORMATION:  
; APPLICANT: TOUTATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,781  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/470,950  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-470-950-4  
Query Match 14.8%; Score 73; DB 1; Length 2101;  
Best Local Similarity 25.0%; Pred. No. 10;  
Matches 21; Conservative 19; Mismatches 32; Indels 12; Gaps 2;  
QY 10 TLQGEAGNFRISGDLKTQIDQVESTAGSLQGWGRGAAGTAQAQAAVYRFOEANKQKQEL 69  
|| ||| : | : : : : | : : | : | : || : || :  
Db 417 TLKQEAATLAANTQIARVEMLETERGQGEAKLLAERG-----HFEE-----EKQQL 464  
QY 70 DEISTNIRQAGVQYSRADDEQQA 93  
: | : : : | : | : || : || :  
Db 465 SSLITDLOSSISNLSQAKELEQA 488  
US-08-467-781-4  
RESULT 11  
US-08-467-781-4  
; Sequence 4, Application US/08467781  
; Patent No. 5780596  
; GENERAL INFORMATION:  
; APPLICANT: TOUTATLY, GARY  
; APPLICANT: LIDGARD, GRAHAM P  
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
; STREET: 125 HIGH STREET  
; CITY: BOSTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,781  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PITCHER ESQ, EDMUND R  
; REGISTRATION NUMBER: 27,829  
; REFERENCE/DOCKET NUMBER: MTP-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2101 amino acids  
; TYPE: amino acid

APPLICANT: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
TITLE OF INVENTION: 6  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA HURWITZ & THIBEAULT  
STREET: 53 STATE STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02109

Search completed: July 5, 2001, 11:46:57  
Job time: 311 sec

Query Match	14.8%	Score 73:	DB 5:	Length 2101:
Best Local Similarity	25.0%	Pred. No. 10:		
Matches 21: Conservative	19:	Mismatches	32:	Indels 12:
				Gaps 2:

RESULT 15  
5510466-2  
; Patent No. 5510466  
; APPLICANT: KREIGER, MONTY; KODAMA, TATSUHIKO  
; TITLE OF INVENTION: SCAVENGER RECEPTOR PROTEIN AND ANTIBODY

Query Match	13.9%	Score 68.5;	DB 6;	Length 349;
Best Local Similarity	26.5%;	Pred. No. 3.3;		
Matches 22;	Conservative 14;	Mismatches 36;	Indels 11;	Gaps 2;





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:42:21 ; Search time 79.63 seconds  
(without alignments)  
95.661 Million cell updates/sec

Title: US-09-462-480-5  
Perfect score: 492  
Sequence: 1 MAEMKTDAAATLGOEAGNFER.....VQYSRADDEQQALSSQMGF 100  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	486	98.8	100	2 H70802	hypothetical prote
2	187	38.0	100	2 T10032	hypothetical prote
3	92.5	18.8	1186	2 T35661	probable chromosom
4	91.5	18.6	410	2 B59103	hypothetical prote
5	90	18.3	750	2 T38435	coiled coil protei
6	89	18.1	103	2 T70600	hypothetical prote
7	87.5	17.8	97	2 E70836	probable transcrip
8	87	17.7	2022	2 T43214	ovtl protein - nam
9	84.5	17.2	97	2 G70857	probable PE protei
10	81.5	16.6	955	2 A47334	Lckin kinesin-rela
11	81	16.5	527	2 S33068	myosin heavy chain
12	81	16.5	1940	2 A59287	myosin heavy chain
13	79.5	16.2	163	2 F75559	hypothetical prote
14	79.5	16.2	1961	1 A61231	myosin heavy chain
15	79.5	16.2	1999	1 S21801	myosin heavy chain
16	79	16.1	245	2 T44704	hypothetical prote
17	79	16.1	508	1 KRSHL2	keratin type II, m
18	77	15.7	561	2 G83378	probable chemotaxi
19	76.5	15.5	528	2 B75116	conserved hypothet
20	76	15.4	491	2 S05408	keratin, type II,
21	75	15.2	2094	2 S33124	tpv protein - huma
22	74.5	15.1	80	2 D29674	phycocyanin linker
23	74.5	15.1	316	2 JC6549	apolipoprotein E p
24	74.5	15.1	387	2 S57834	fcrA protein precu
25	74.5	15.1	880	2 F83386	hypothetical prote
26	74.5	15.1	929	2 F51932	kinesin [imported]
27	74.5	15.1	1313	2 F96673	hypothetical prote
28	74.5	15.1	1956	2 T16416	hypothetical prote
29	74	15.0	756	2 E75590	methyl-accepting c

30	74	15.0	1325	2 T42722	male-enhanced anti
31	74	15.0	2241	2 T20971	hypothetical prote
32	74	15.0	2261	2 T20978	hypothetical prote
33	73.5	14.9	388	2 S25336	fcrA 15 protein -
34	73.5	14.9	433	2 D84335	hypothetical prote
35	73.5	14.9	440	1 S33479	ferredoxin--NADP+
36	73.5	14.9	745	2 H96760	unknown protein (1
37	73.5	14.9	975	1 A31497	kinesin heavy chai
38	73	14.8	257	2 I38025	keratin-like prote
39	73	14.8	860	2 T14650	tail fiber protein
40	73	14.8	978	2 T14968	phage lambda-relat
41	73	14.8	1589	2 T13606	hypothetical prote
42	73	14.8	2101	2 A42184	nuclear mitotic ap
43	72.5	14.7	80	2 A24691	8.9K linker polype
44	72.5	14.7	405	2 A39339	FC gamma (IgG) rec
45	72.5	14.7	576	2 A39228	flagellin A - camp

## ALIGNMENTS

### RESULT 1

H70802  
hypothetical protein RV3874 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: H70802  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: H70802  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-100 <COL>  
A:Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAAL17966.1; PID:g296  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: RV3874

Query Match	98.8%	Score	486	DB	2	Length	100
Best Local Similarity	99.0%	Pred. No.	3.1e-37				
Matches	99	Conservative	0	Mismatches	1	Indels	0
Gaps							

QY	1	MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOE	60
Db	1	MAEMKTDAAATLGOEAGNFERISGDLKTQIDQVESTAGSLQGWGAGTAQAQAAVVRFOE	60

QY	61	AANKQKQELDEISTNRQAGVQYSRADDEQQALSSQMGF	100
Db	61	AANKQKQELDEISTNRQAGVQYSRADDEQQALSSQMGF	100

RESULT 2  
T10032  
hypothetical protein MLCB28.13c - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000  
C:Accession: T10032  
R:Biglmeier, K.; Honore, N.; Woods, S.A.; Caudron, B.; Cole, S.T.  
Mol. Microbiol. 7, 197-206, 1993  
A:Title: Use of an ordered cosmid library to deduce the genomic organization of Mycob  
A:Reference number: Z16917; MUID:93188700  
A:Accession: T10032  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-100 <BIG>  
A:Cross-references: EMBL:Y14967; NID:g2370268; PIDN:CAA75210.1; PID:g2370280

Db 11 QLEQAARTVKNTKSSLEYTHQDLYSOTEXYIAS-----QWGSASSD-----RFYQMF 56

QY 63 NKOK-----QELDEISTNIRQAGVQYSRADE 88  
| : |  
| | | | : | : | : | : |  
Db 57 NEAKPMFNILQELDKIAYELERAANKFREADE 89

RESULT 5  
T38435  
coiled coil protein - fission yeast (Schizosaccharomyces pombe)  
C-Species: Schizosaccharomyces pombe  
C-Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C-Accession: T38435  
R:McDougall, R.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, December 1997  
A-Reference number: Z21793  
A-Accession: T38435  
A-Status: preliminary; translated from GB/EMBL/DDBJ  
A-Molecule type: DNA  
A-Residues: 1-750 <MCD>  
A-Cross-references: EMBL:AL009227; PIDN:CAAL5821.1; GSPDB:GN00066; SPDB:SPAC27D7.02C  
A-Experimental source: strain 972h-; cosmid c27D7  
C-Genetics:  
A-Gene: SPDB:SPAC27D7.02C  
A-Map position: 1

Db	550	KQAGENHYNLSDYETQIKSLESL-----TNSQAECYSFQEKINELNSQID	597
QY	71	EISTNIRQAGVQYSRADDEEQQAALS	95
Db	598	ELKLLNEANKKY-----QELAIS	616

C:Accession: B70600  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: B70600  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-103 <COL>  
 A:Cross-references: GB:T94121; GB:Al123456; NID:g3261736; PIDN:CAB08096.1; PTD:e31227  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV3905C

QY	4	MKTD	A	T	L	G	E	A	G	N	F	R	I	S	D	L	K	T	O	I	D	O	V	E	S	T	A	G	S	L	Q	G	O	W	R	G	A	A	G	T	A	A	O	A	A	V	V	F	O	E	A	N	63				
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
Db	7	LR	V	E	P	A	V	M	Q	G	F	A	A	L	D	G	A	E	H	L	A	V	O	L	A	E	L	D	A	O	V	Q	M	L	G	O	W	R	G	A	S	G	A	S	A	W	E	L	H	R	G	A	66				
QY	64	KQ	E	L	E	I	S	T	N	I	R	O	A	G	V	O	Y	S	R	A	E	E	Q	O	A	L	94																														

Db 67 EVGLGSLMLAAIAHAGAGVQHNETASQVL 97

## RESULT 7

E70836 probable transcription regulator RV0287 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: E70836

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987

A: Accession: E70836

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-97 <COL>

A: Cross-references: GB:AL021930; GB:AL123456; NID: g3261524; PIDN: CAA17362.1; PID: el25247

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: RV0287

Query Match 17.8%; Score 87.5; DB 2; Length 97;

Best Local Similarity 34.9%; Pred. No. 0.24; Mismatches 43; Indels 3; Gaps 1;

Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

QY 18 FERISGLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQVRFQEAANKQKQELDEISTNIR 77

Db 18 FAKAGLMRHTTICQAEQQAASQAQAFHOGESSAFAQAHARFVAAAKVNTLLDVAQANLG 77

QY 78 QAGVQYSRADEEQQAALSSOMGF 100

Db 78 EAAGTYVAAD---AAAASVTGTF 97

## RESULT 8

T43214

ovt1 protein - nematode (Onchocerca volvulus)

N: Alternate names: myosin-like antigen

C: Species: Onchocerca volvulus

C: Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 20-Jun-2000

C: Accession: T43214; A44939; A54513; S27825

R: Friteeraprab, S.; Richie, T.L.; Tuan, R.S.; Shepley, K.J.; Dinman, J.D.; Neubert, T. Mol. Biochem. Parasitol. 69, 161-171, 1995

A: Title: Molecular cloning of a gene expressed during early embryonic development in On

A: Reference number: 222341; MUID: 95287898

A: Accession: T43214

A: Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: mRNA

A: Residues: 1-2022 <TRI>

A: Cross-references: EMBL: U12681; NID: g530824; PID: g530825; PIDN: AAA80009.1

A: Experimental source: specific host Homo sapiens

R: Erond, N.E.; Donelson, J.E.

Mol. Biochem. Parasitol. 40, 213-224, 1990

A: Title: Characterization of a myosin-like antigen from Onchocerca volvulus.

A: Reference number: A44939; MUID: 90301142

A: Accession: A44939

A: Molecule type: mRNA

A: Residues: 733-874, 'E', 876-916, 'S', 918-1038, 1040-1047, 'S', 1049-1283, 'E', 1285-1363 <ERO>

A: Cross-references: GB: M30398

A: Note: the sequence is revised in Genbank entry ONGANTML, release 115, (PIDN: AAA29413.1)

R: Donelson, J.E.; Duke, B.O.L.; Moser, D.; Zeng, W.; Erond, N.E.; Lucius, R.; Renz, A.

Mol. Biochem. Parasitol. 31, 241-250, 1988

A: Title: Construction of Onchocerca volvulus cDNA libraries and partial characterization

A: Reference number: A54513; MUID: 89127417

A: Accession: A54513

A: Status: preliminary

A: Molecule type: mRNA

A: Residues: 733-866 <DON>

A: Cross-references: GB: J03995; NID: g159874; PIDN: AAA29412.1; PID: g159875

R: Ngozi, E.; Erond, N.E.; Donelson, J.E.

submitted to the EMBL Data Library, April 1990

A: Description: Characterization of a myosin-like antigen from Onchocerca volvulus.

A: Reference number: S27825

A: Molecule type: mRNA

A: Residues: 733-874, 'E', 876-916, 'S', 918-1038, 1040-1047, 'S', 1049-1283, 'E', 1285-1347, 'E'

A: Cross-references: EMBL: M30398; NID: g159876; PIDN: AAA29413.1; PID: g159877

C: Genetics:

A: Gene: ovt1

C: Keywords: leucine zipper

Query Match 17.7%; Score 87; DB 2; Length 2022;

Best Local Similarity 25.5%; Pred. No. 7.5; Mismatches 41; Indels 10; Gaps

Matches 25; Conservative 22; Mismatches 41; Indels 10; Gaps

QY 11 LGQEAAGNFERISGDLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQVRFQEAANKQKQELDEISTNIR 98

Db 1752 LDEKRTMENILHETALQREAISSNLALERENKELHRNCAQLQQIAOLELENGRLIQ 1811

QY 61 AANKQKQELDEISTNIRQAGVQYSRADEEQQAALSSOM 98

Db 1812 LTNKQREYDKFAQNMRTKTIQIETIENRERSLSKRI 1849

## RESULT 9

G70857

probable PE protein - Mycobacterium tuberculosis (strain H37RV)

C: Species: Mycobacterium tuberculosis

C: Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C: Accession: G70857

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A: Reference number: A70500; MUID: 98295987

A: Accession: G70857

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-97 <COL>

A: Cross-references: GB: AL021287; GB: AL123456; NID: g3261508; PIDN: CAA16105.1; e123

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PE

Query Match 17.2%; Score 84.5; DB 2; Length 97;

Best Local Similarity 34.9%; Pred. No. 0.45; Mismatches 29; Conservative 7; Mismatches 44; Indels 3; Gaps 1;

Matches 29; Conservative 7; Mismatches 44; Indels 3; Gaps 1;

QY 18 FERISGLTKTQIDQVESTAGSLQGWGAGTAAGTAQAQVRFQEAANKQKQELDEISTNIR 77

Db 18 FAKAGLMRHTTICQAEQQAASQAQAFHOGESSAFAQAHARFVAAAKVNTLLDVAQANLG 77

QY 78 QAGVQYSRADEEQQAALSSOMGF 100

Db 78 EAAGTYVAAD---AAAASVTGTF 97

## RESULT 10

A47334

Lckin kinesin-related antigen - Leishmania chagasi (fragment)

C: Species: Leishmania chagasi

C: Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001

C: Accession: A47334

R: Burns Jr., J.M.; Shreffler, W.G.; Benson, D.R.; Ghalib, H.W.; Badaro, R.; Reed, S.G.

Proc. Natl. Acad. Sci. U.S.A. 90, 775-779, 1993

A: Title: Molecular characterization of a kinesin-related antigen of Leishmania chagasi

A;Reference number: A47334; MUID:93133867

A;Accession: A47334

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-955 <BUR>

A;Cross-references: GB:107879; NID:g308884; PIDN:AAA29254.1; PID:g308885

A;Experimental source: MHOM/BR/82/BA-2,C1

A;Note: sequence extracted from NCBI backbone (NCBIN:122864, NCBI:P:122865)

C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C;Keywords: ATP; nucleotide binding; P-loop

F;13-398/Domain: kinesin motor domain homology <KNOT>

F;122-129/Region: nucleotide-binding motif A (P-loop)

Query Match 16.5%; Score 81.5; DB 2; Length 955;

Best Local Similarity 29.2%; Pred. No. 10;

Matches 31; Conservative 15; Mismatches 47; Indels 13; Gaps 2;

Qy 3 EMKTDATLGOEAGNFERISGLTKTQIDQV-----ESTAGSLQGWGAACTAA 51

Db 852 DRESTRATLEQQLRESEERAAELASQLESTAAKMSAEQDRESTRATLEQQLRDSERAA 911

Qy 52 QAAVVFQEAANKQKQELDEISTNIRQAGVQYSDAEQQAALSSQ 97

Db 912 ELASQLEATAAKSSAEQDR--ENTRALEQQLRDSERAAELASQ 955

RESULT 11

S33068 myosin heavy chain - fluke (Schistosoma mansoni) (fragment)

N;Alternate names: surface antigen, 200K

C;Species: Schistosoma mansoni

C;Date: 22-Nov-1993 #sequence\_revision 06-Sep-1996 #text\_change 13-Feb-1998

C;Accession: S33068

J;Soisson, L.M.A.; Masterson, C.P.; Tom, T.D.; McNally, M.T.; Lowell, G.H.; Strand, M.

J;Immunol 149, 3612-3620, 1992

A;Title: Induction of protective immunity in mice using a 62-kDa recombinant fragment of

A;Reference number: A46514; MUID:93056536

A;Accession: S33068

A;Molecule type: mRNA

A;Residues: 1-527 <SOI>

A;Cross-references: EMBL:X65591

A;Note: the authors translated the codon CAA for residue 346 as Lys

C;Superfamily: myosin heavy chain; myosin motor domain homology

C;Keywords: ATP; surface antigen

Query Match 16.5%; Score 81; DB 2; Length 527;

Best Local Similarity 21.9%; Pred. No. 6;

Matches 21; Conservative 24; Mismatches 45; Indels 6; Gaps 2;

Qy 3 EMKTDATLGOEAGNFERISGLTKTQIDQVSTAGSLQGWGAACTAAQAAVVRFOEAA 62

Db 337 KLEGDLKATQETVDDLERVKRDLERKRAEIGGLSGKFEDEQGLVAQ-----LORKI 391

Qy 63 NKQKQELDEISTNIRQAGVQYSDAEQQAALSSQ 98

Db 392 KELQTRIQLLEEDLEAERAARSKAERSQ--LESEL 426

RESULT 12

A59287

myosin heavy chain - fluke (Schistosoma mansoni) (strain Brazilian LE)

C;Species: Schistosoma mansoni

C;Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 08-Sep-2000

C;Accession: A59287

R;Weston, D.S.; Schmitz, J.; Kemp, M.; Kunz, W.

Mol. Biochem. Parasitol. 58, 161-164, 1993

A;Title: Cloning and sequence characterization of a complete myosin heavy chain cDNA frq

A;Reference number: A59287; MUID:93211444

A;Accession: A59287

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-1940 <WES>

A;Cross-references: GB:L01634; PIDN:AAA29905.1

A;Experimental source: strain Brazilian LE

C;Genetics:

A;Gene: MYH

C;Superfamily: myosin heavy chain; myosin motor domain homology

F;82-752/Domain: myosin motor domain homology <MMO>

Query Match 16.5%; Score 81; DB 2; Length 1940;

Best Local Similarity 21.9%; Pred. No. 25;

Matches 21; Conservative 24; Mismatches 45; Indels 6; Gaps 2;

Qy 3 EMKTDATLGOEAGNFERISGLTKTQIDQVSTAGSLQGWGAACTAAQAAVVRFOEAA 62

Db 1040 KLEGDLKATQETVDDLERVKRDLERKRAEIGGLSGKFEDEQGLVAQ-----LORKI 1094

Qy 63 NKQKQELDEISTNIRQAGVQYSDAEQQAALSSQ 98

Db 1095 KELQTRIQLLEEDLEAERAARSKAERSQ--LESEL 1129

RESULT 13

F75559

hypothetical protein - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C;Accession: F75559

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896

A;Accession: F75559

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-163 <WHI>

A;Cross-references: GB:AE001873; GB:AE000513; NID:g6457764; PIDN:AAF09701.1; PID:g645

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR0105

A;Map position: 1

Query Match 16.2%; Score 79.5; DB 2; Length 163;

Best Local Similarity 22.7%; Pred. No. 2.3;

Matches 22; Conservative 16; Mismatches 46; Indels 13; Gaps 1;

Qy 11 LGQEAGN-----FERISGLTKTQIDQVSTAGSLQGWGAACTAAQAAVVR 57

Db 1 MGDEAGNAQDRAEAAAAARAARAEASHGADYRRREMSANDDLLSLGLQAKHKTDQAAAEKVRAD 60

Qy 58 FOEAANKQKQELDEISTNIRQAGVQYSDAEQQAAL 94

Db 61 TOEAQNAREKAQDVVRANVHESAQDFRAGAEQAQTL 97

RESULT 14

A61231

myosin heavy chain nonmuscle form A - human

N;Alternate names: cellular myosin heavy chain; myosin type 9; NMHC-A

N;Contains: myosin ATPase (EC 3.6.1.32)

C;Species: Homo sapiens (man)

C;Date: 12-May-1994 #sequence\_revision 14-Jul-1994 #text\_change 19-Jan-2001

C;Accession: A61231; A34876; I52562; I61692

R;Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelste

Circ. Res. 69, 530-539, 1991

A;Title: Human nonmuscle myosin heavy chains are encoded by two genes located on diff

A;Reference number: A61231; MUID:91316803

A;Accession: A61231

A;Molecule type: mRNA

A;Residues: 1-715 <SIM>

A:Cross-references: GB:M69180; NID:q189029; PIDN:AAAG1765.1; PID:q189030  
R:Saiz, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.  
Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990  
A:Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through altern  
A:Reference number: A34876; MUID:90138958  
A:Accession: A34876  
A:Molecule type: mRNA  
A:Residues: 715-1961 <SAE>  
A:Cross-references: GB:M31013; NID:q189035; PIDN:AAA36349.1; PID:q189036  
R:Toothaker, L.E.; Gonzalez, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.; Arnaout, M.A.;  
Blood 78, 1826-1833, 1991  
A:Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones,  
A:Reference number: I52562; MUID:92003925  
A:Accession: I52562  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-52, 'ENI', 56-659, 'T', 661-868, 'T', 870-930, 'C', 932-1239, 'KG', 1242-1337 <RES>  
A:Cross-references: GB:M81105; NID:q189888; PIDN:AAA59888.1; PID:q553596  
R:Bement, W.M.; Hasson, T.; Wirth, J.A.; Cheney, R.E.; Mooseker, M.S.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994  
A:Title: Identification and overlapping expression of multiple unconventional myosin gen  
A:Reference number: A55758; MUID:94294418  
A:Accession: I61692  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 182-218 <BBM>  
A:Cross-references: GB:L29141; NID:q457249; PIDN:AAA20904.1; PID:q531134  
C:Genetics:  
A:Gene: GDB:MYH9  
A:Cross-references: GDB:120216; OMIM:160775  
A:Map position: 22q12.3-22q13.1  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid  
F:84-764/Domain: myosin motor domain homology <MMOT>  
F:174-181/Region: nucleotide-binding motif A (P-loop)  
F:552-565/Region: actin binding #status predicted  
F:626-640/Region: actin binding #status predicted  
F:837-1938/Domain: coiled coil #status predicted <COI>  
F:837-1277/Domain: S2 #status predicted <DS2>  
F:1278-1961/Domain: light meromyosin #status predicted <LMW>  
F:1939-1961/Domain: carboxyl-terminal <CBT>  
F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
F:180/Binding site: ATP (Lys) #status predicted  
F:694,704/Active site: Cys #status predicted

Query Match 16.2%; Score 79.5; DB 1; Length 1961;  
Best Local Similarity 23.1%; Pred. No. 35;  
Matches 31; Conservative 26; Mismatches 28; Indels 49; Gaps 5;  
QY 9 ATLGQAGN-----PERISGDLKTDQVSTAGSLQGWGAGTAQAQAVVRF 58  
DB 1765 ADLNLGRHQAQKNENARQQLERQNKELVKLOEME---GTVSKYK-ASITALEAKIAQL 1820  
QY 59 QE-----AANKQ-----KOELDEISTNIRQAGVQY 83  
DB 1821 EEQLDNETKERQAQKQVRTEKKLDVLLQVDDERRNAEQYKQADKASTRLKQLKROL 1880  
QY 84 SRADEEQQALSSQ 97  
DB 1881 EEAEERAAQANASR 1894

RESULT 15  
myosin heavy chain, neuronal [similarity] - rat  
N:Alternate names: myosin II  
N:Contains: myosin ATPase (EC 3.6.1.32)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Dec-1993 #sequence-revision 31-Dec-1993 #text\_change 19-Jan-2001  
C:Accession: S21801; PNO013; S18134  
R:Sun, W.; Chantler, P.D.  
J. Mol. Biol. 224, 1185-1193, 1992

A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian br  
A:Reference number: S21801; MUID:92235856  
A:Accession: S21801  
A:Molecule type: mRNA  
A:Residues: 1-1999 <SUN>  
A:Cross-references: EMBL:X62659  
R:Sun, W.; Chantler, P.D.  
Biochem. Biophys. Res. Commun. 175, 244-249, 1991  
A:Title: A unique cellular myosin II exhibiting differential expression in the cerebr  
A:Reference number: PNO013; MUID:91151356  
A:Accession: PNO013  
A:Molecule type: mRNA  
A:Residues: 1914-1998, 'I' <SU2>  
A:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleo  
F:84-763/Domain: myosin motor domain homology <MMOT>  
F:174-181/Region: nucleotide-binding motif A (P-loop)  
F:541-575/Region: actin binding #status predicted  
F:653-675/Region: actin binding #status predicted  
F:836-1999/Domain: coiled coil #status predicted <COI>  
F:836-1276/Region: S2  
F:1277-1999/Region: light meromyosin  
F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted  
F:180/Binding site: ATP (Lys) #status predicted  
F:693,703/Active site: Cys #status predicted  
F:1916/Binding site: phosphate (Ser) (covalent) #status predicted  
F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 16.2%; Score 79.5; DB 1; Length 1999;  
Best Local Similarity 23.1%; Pred. No. 35;  
Matches 31; Conservative 26; Mismatches 28; Indels 49; Gaps 5;  
QY 9 ATLGQAGN-----PERISGDLKTDQVSTAGSLQGWGAGTAQAQAVVRF 58  
DB 1764 ADLNLGRHQAQKNENARQQLERQNKELVKLOEME---GTVSKYK-ASITALEAKIAQL 1819  
QY 59 QE-----AANKQ-----KOELDEISTNIRQAGVQY 83  
DB 1820 EEQLDNETKERQAQKQVRTEKKLDVLLQVDDERRNAEQYKQADKASTRLKQLKROL 1879  
QY 84 SRADEEQQALSSQ 97  
DB 1880 EEAEERAAQANASR 1893

Search completed: July 5, 2001, 11:48:30  
Job time: 369 sec

7

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:43:21 ; Search time 41.8 Seconds  
(without alignments)  
81.951 Million cell updates/sec

Title: US-09-462-480-5  
Perfect score: 492  
Sequence: 1 MAEMKTDAAATLQGEAGNFER.....VOYSGRADEQQALSSQMGF 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	87	17.7	2022	1 ANTL_ONCVO	P21249 onchocerca
2	81.5	16.6	955	1 KINL_LEICH	P46865 leishmania
3	78.5	16.0	1960	1 MYSN_HUMAN	P35579 homo sapien
4	76	15.4	491	1 K2M2_SHEEP	P15241 ovis aries
5	75.5	15.3	440	1 FENR_ANAVA	Q44549 anabaena va
6	75	15.2	2349	1 TPR_HUMAN	P12270 homo sapien
7	74.5	15.1	80	1 PYSL_ANASP	P07124 anabaena sp
8	74	15.0	1325	1 G160_MOUSE	P55937 mus musculus
9	73.5	14.9	440	1 FENR_ANASO	P21890 anabaena sp
10	73.5	14.9	975	1 KINH_DROME	P17210 drosophila
11	73	14.8	1589	1 PHP_DROME	P39769 drosophila
12	72.5	14.7	80	1 PYSL_NASLA	P11396 mastigoclad
13	72.5	14.7	575	1 FLA2_CAMJE	P22251 campylobact
14	72.5	14.7	1679	1 Y109_YEAST	P40457 saccharomyc
15	72	14.6	502	1 K2M3_SHEEP	P25691 ovis aries
16	71.5	14.5	292	1 BPHC_PSES1	P17297 pseudomonas
17	71.5	14.5	1959	1 MYSN_CHICK	P14105 gallus gall
18	71	14.4	457	1 MESE_LEUME	Q10419 leuconostoc
19	71	14.4	548	1 HLYB_VIBCH	P15492 vibrio chol
20	70.5	14.3	78	1 PYSL_SYNEL	P50035 synechococc
21	70.5	14.3	845	1 Y4FA_RHINL	P55439 rhizobium s
22	70	14.2	284	1 T2M2_SCHMA	P42638 schistosoma
23	70	14.2	512	1 MCPD_ENTAE	P21823 enterobacte
24	70	14.2	535	1 HTRL_HALHA	P33741 halobacteri
25	70	14.2	535	1 HTRL_HALNI	Q9hpf6 halobacteri
26	70	14.2	576	1 PEX5_PICPA	P33292 picchia past
27	70	14.2	715	1 CLPB_MYCPN	P75247 mycoplasma
28	69.5	14.1	1172	1 LMB3_HUMAN	Q13751 homo sapien
29	69	14.0	186	1 APL3_GALME	P80703 gallieria me
30	69	14.0	189	1 APL3_MANSE	P13276 manduca sex
31	69	14.0	1938	1 MYSD_CAREL	P02567 caenorhabdi
32	68.5	13.9	356	1 IRPA_SYNP7	P21608 synechococc
33	68.5	13.9	453	1 MSRE_BOVIN	P21758 bos taurus

34 68.5 13.9 1107 1 YJEP\_ECOLI P39285 escherichia  
35 68.5 13.9 1947 1 MYSC\_CAREL P12845 caenorhabdi  
36 68 13.8 213 1 MSRA\_ERWCH Q92eq8 erwinia chr  
37 68 13.8 384 1 K2CD\_HUMAN P48667 homo sapien  
38 68 13.8 563 1 K2CE\_HUMAN P48668 homo sapien  
39 68 13.8 1937 1 MYSP\_HUMAN P13535 homo sapien  
40 67.5 13.7 388 1 MRP4\_STRPY P30141 streptococc  
41 67.5 13.7 590 1 DNAK\_STRMU O06942 streptococc  
42 67.5 13.7 643 1 CTK2\_XENLA P79955 xenopus lae  
43 67.5 13.7 795 1 GCR\_RAT P06536 rattus norv  
44 67.5 13.7 1938 1 MYSS\_CHICK P13538 gallus gall  
45 67 13.6 245 1 YT27\_MYCTU Q10973 mycobacteri

#### ALIGNMENTS

RESULT 1  
ANTL\_ONCVO  
ID ANTL\_ONCVO STANDARD; PRT; 2022 AA.  
AC P21249;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE MAJOR ANTIGEN.  
GN OVT1.  
OS Onchocerca volvulus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
OC Onchocercidae; Onchocerca.  
OX NCBI\_TaxID=6282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95287898; PubMed=7770081;  
RA Tricetrappap S., Richie T.L., Tuan R.S., Shepley K.J., Dinman J.D.,  
RA Neubert T.A., Scott A.L.;  
RT "Molecular cloning of a gene expressed during early embryonic  
RT development in Onchocerca volvulus."  
RL Mol. Biochem. Parasitol. 69:161-171(1995).  
RN [2]  
RP SEQUENCE OF 733-866 FROM N.A.  
RX MEDLINE=89127417; PubMed=2464764;  
RA Donelson J.E., Duke B.O.L., Moser D., Zeng W., Erundu N.E.,  
RA Lucius R., Renz A., Karam M., Flores G.Z.;  
RT "Construction of Onchocerca volvulus cDNA libraries and partial  
RT characterization of the cDNA for a major antigen."  
RL Mol. Biochem. Parasitol. 31:241-250(1988).  
CC -|- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: U12681; AAA80009.1; -  
DR EMBL: J03995; AAA29412.1; -  
DR PIR: A54513; A54513.  
DR HSSP: P02633; 3ICB.  
KW Antigen; Coiled coil.  
FT DOMAIN 74 120 COILED COIL (POTENTIAL).  
FT DOMAIN 151 251 COILED COIL (POTENTIAL).  
FT DOMAIN 327 384 COILED COIL (POTENTIAL).  
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).  
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;

Query Match 17.7%; Score 87; DB 1; Length 2022;

Best Local Similarity 25.5%; Pred. No. 3.2;  
Matches 25; Conservative 22; Mismatches 41; Indels 10; Gaps 2;

QY 11 LGQAGNFRISGDLTKTDQVESTAGSLQGG-----WRGAAGTAQAQAVV-----RFQE 60





```
DR HSP: P08799; 1MND.
DR MIM; 160775; -.
DR InterPro: IPR000048; -.
DR InterPro: IPR001609; -.
DR InterPro: IPR002928; -.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF01576; myosin_tail; 1.
DR Pfam: PF00063; myosin_head; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PROSITE: PS00996; IQ; 1.
KW Myosin; coiled coil; Actin-binding; Alkylation; ATP-binding;
KW Multigene family; Calmodulin-binding.
FT DOMAIN 1 836 GLOBULAR HEAD (S1).
FT DOMAIN 837 1960 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 837 1960 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP.
FT DOMAIN 654 676 ACTIN-BINDING.
FT DOMAIN 786 800 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 812 829 CALMODULIN-BINDING (BY SIMILARITY).
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 53 55 EAI -> RGH (IN REF. 3).
FT CONFLICT 660 660 T -> S (IN REF. 3).
FT CONFLICT 869 869 T -> M (IN REF. 4).
FT CONFLICT 931 931 C -> Y (IN REF. 4).
FT CONFLICT 1240 1241 KG -> GR (IN REF. 4).
FT CONFLICT 1350 1350 E -> EE (IN REF. 2).
FT CONFLICT 1764 1764 T -> A (IN REF. 2).
FT CONFLICT 1771 1771 S -> G (IN REF. 2).
SQ SEQUENCE 1960 AA; 226531 MW; 588F84BB8C106E6F CRC64;

Query Match 16.0%; Score 78.5; DB 1; Length 1960;
Best Local Similarity 21.7%; Pred. No. 18;
Matches 34; Conservative 23; Mismatches 29; Indels 71; Gaps 5;

QY 11 LQEAQNGFERISGLK---TQIDQV-----EST 35
DB 1738 LEEEOGNTLNDRLKANQIQINTDLNLSHAQKNENARQQLERQNKELVKLQEM 1797
QY 36 AGSLOGWRGAAGTAQAQAAVVFQEE-----RANKQ----- 65
DB 1798 EGTVKSKYK-ASITALEAKTAQLEQDNETKQRAQKQVRTEKKLQVLLQVDDERR 1856
QY 66 -----KQELDEISTNIRAGQVYSRADEEQQALSSQ 97
DB 1857 NAEQYKDQADKASTRLKQLKRLQLEAEAEAEQAQANASR 1893

RESULT 4
K2M2_SHEEP
ID K2M2_SHEEP STANDARD; PRT; 491 AA.
AC P15241.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE KERATIN, TYPE II MICROFIBRILLAR, COMPONENT 7C.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RX MEDLINE=90026244; PubMed=2803231;
RA Sparrow L.G., Robinson C.P., McMahon D.T.W., Rubira M.R.;
RT "The amino acid sequence of component 7c, a type II intermediate-
RT filament protein from wool."
RL Biochem. J. 261:1015-1022(1989).
CC -!- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
```

```
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -!- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,
CC 7A, 7B, AND 7C).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR PIR: S05408; S05408.
DR InterPro: IPR001664; -.
DR InterPro: IPR003054; -.
DR Pfam: PF00038; filament; 1.
DR PRINTS: PR01276; TYPE2KERATIN.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin.
FT MOD_RES 1 1 BLOCKED.
FT DOMAIN 1 109 HEAD.
FT DOMAIN 110 416 ROD.
FT DOMAIN 417 491 TAIL.
FT DOMAIN 110 144 COIL 1A.
FT DOMAIN 145 154 LINKER 1.
FT DOMAIN 155 255 COIL 1B.
FT DOMAIN 256 272 LINKER 12.
FT DOMAIN 273 416 COIL 2.
FT VARIANT 74 74 C -> G OR S.
FT VARIANT 80 80 C -> S.
FT VARIANT 144 144 F -> Y.
FT VARIANT 232 232 S -> V.
FT VARIANT 276 276 C -> D OR N.
FT VARIANT 284 284 Q -> H.
FT UNSURE 1 2 CG -> GC.
SQ SEQUENCE 491 AA; 53681 MW; A80171FE3831ABE CRC64;

Query Match 15.4%; Score 76; DB 1; Length 491;
Best Local Similarity 22.9%; Pred. No. 6.9;
Matches 19; Conservative 23; Mismatches 39; Indels 2; Gaps 1;

QY 17 NFERISGDKLTQIDQVESTAGSLQGWGARGAAGTAQAQAAVVFQEAANKQKQELDEISTNI 76
DB 273 NMDCIVAEIKAIQYDDIASRAEASWYSKCEIKATVIRHGETLRRTRKEEINELNRI 332
QY 77 RQ--AGVQYSRADEEQQALSSQ 97
DB 333 QRLTAEVENAKQNSKLEAAVTQ 355

RESULT 5
FENR_ANAVA
ID FENR_ANAVA STANDARD; PRT; 440 AA.
AC Q44549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FERREDOXIN--NADP REDUCTASE (EC 1.18.1.2) (FNR).
GN PETH.
OS Anabaena variabilis.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=1172;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29413.
RA Mannan R.M., Matthijs H.C.P., Pakrasi H.B.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: REDUCED FERREDOXIN + NADP(+) = OXIDIZED
CC FERREDOXIN + NADPH.
CC -!- COFACTOR: FAD.
CC -!- SUBCELLULAR LOCATION: MAY BE BOUND TO THE THYLAKOID MEMBRANE OR
CC ANCHORED TO THE THYLAKOID-BOUND PHYCOBILISOMES.
CC -!- SIMILARITY: IN THE N-TERMINAL REGION, TO PHYCOBILISOME LINKER.
CC PROTEIN CPCD.
CC -!- SIMILARITY: WITH OTHER SPECIES FNR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```



```

Query Match      15.0%; Score 74; DB 1; Length 1325;
Best Local Similarity 24.5%; Pred. No. 30;
Matches 27; Conservative 23; Mismatches 34; Indels 26; Gaps
QY 9 ATLLGQAGNFRFISGDKLTQIDQVESTAG-----SLQGNWRGAAG 48
      : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 AAAAQHQDQNGAEVRRSRDSDICSSVMSSIAQDELLQIKDKRRLEQVLSL 260

```

```
CC
-----
EMBL: X05239; CAA38865.1; -.
DR DR
PIR: D29674; D29674. -.
DR DR
InterPro: IPR001685; -.
DR DR
Pfam: PF01383; Cpcd; 1.
DR DR
PhycoBLISSome; Photosynthesis.
SQ SQ
SEQUENCE 80 AA; 8895 MW; 52FD2D7DBEF5444D5 CRC64;
CC
```

ID	AC	FENR_ANASO	STANDARD;	FRI;	440 AA.
P21890;					
DT	01-MAY-1991	(Rel. 18, Created)			
DT	01-OCT-1994	(Rel. 30, Last sequence update)			

DE FERREDOXIN--NADP REDUCTASE (EC 1.18.1.2) (FNR).  
GN PETH.  
OS *Anabaena* sp. (strain PCC 7119).

NCBI\_PatID=1168;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE-93344523; PubMed-8343609;

RT "Homology of the N-terminal domain of the pETH gene product from  
RT *Arabidopsis* sp. PCC 7119 to the CopD phycobilisome linker polypeptide.";  
RL *Plant Mol. Biol.* 22:725-729(1993).  
RN [2]

SEQUENCE OF 137-440 FROM N.A.  
RX MEDLINE-9108322; PubMed-2124680;  
RA Fillat M.F., Bakker H.A.C., Weisbeek P.J.;  
RT "sequence of the ferredoxin-NADP(+) reductase gene from *Anabaena PCC*  
7119";

RL NUCLEIC ACIDS RES. 18:161-161(1990).  
 RN [3]  
 RP SEQUENCE OF 152-183.  
 RC STRAIN=1403.46;

Sancho J., Peleato M.L., Gomez-Moreno C., Edmondson D.E.;  
 "Purification and properties of ferredoxin-NADP+ oxidoreductase from  
 the nitrogen-fixing cyanobacteria *Anabaena variabilis*.";

RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 137-440.  
 RX MEDLINE-97045988; PubMed-8890910;

RA Pontecilla-Camps J.C., Frey M.;  
RT "X-ray structure of the ferredoxin:NADP+ reductase from the  
RT cyanobacterium *Anabaena PCC 7119* at 1.8-Å resolution, and  
RA Seire D., Verilleux F.M.D., Medina M., Gomez-Moreno C.,  
RA Pontecilla-Camps J.C., Frey M.;

RL J. Mol. Biol. 263:20-39(1996).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 146-440.

RA MELLONE-5511350; FUMOSO-5242134;  
RA Martinez-Julvez M., Hermoso J., Hurley J.K., Mayoral T.,  
RA Sanz-Aparicio J., Tollin G., Gomez-Moreno C., Medina M.;  
RA "Role of Arg100 and Arg264 from Anabaena PCC 7119 ferredoxin-NADP+  
RT reductase for optimal NADP+ binding and electron transfer.";   
RT *Biochemistry* 37:17680-17691(1998).

CC -1- CATALYTIC ACTIVITY: REDUCED FERREDOXIN + NADP(+) = OXIDIZED  
CC FERREDOXIN + NADPH.  
CC -1- COFACTOR: FAD.  
CC -1- SUBCELLULAR LOCATION: MAY BE BOUND TO THE THYLAKOID MEMBRANE OR  
CC ANCHORED TO THE THYLAKOID-BOUND PHYCOBILISOMES.  
CC -1- SIMILARITY: WITH OTHER SPECIES FNR.  
CC -1- SIMILARITY: IN THE N-TERMINAL REGION, TO PHYCOBILISOME LINKER

CC  
PROTEIN CPD.  
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC DR EMBL; X72394; CAA51088.1; -.
CC DR EMBL; X54039; CAA37973.1; -.
CC DR PIR; A27581; A27581.
CC DR PIR; S13103; S13103.
CC DR PIR; S33479; S33479.
CC DR PIR; S35150; S35150.
CC DR PDB; 1QUE; 15-MAY-97.
CC DR PDB; 1QUF; 17-SEP-97.
CC DR PDB; 1BJK; 18-NOV-98.
CC DR InterPro; IPR001433; -.
CC DR InterPro; IPR001685; -.
CC DR InterPro; IPR001709; -.
CC DR Pfam; PF00175; cxdred_fad; 1.
CC DR PRINTS; PR00371; FPCR.
CC DR Oxidoreductase; Flavoprotein; NADP; FAD; Thylakoid membrane;
CC Phycobilisome; 3D-structure.
CC FT DOMAIN 1 80 CPCD-LIKE.
CC FT NP_BIND 288 306 NADP (RIBOSE PART) (BY SIMILARITY).
CC FT CONFLICT 180 180 I -> L (IN REF. 3).
CC FT SEQUENCE 440 AA; 48865 MW; 8E1F61D0F09338B6 CRC64;

Query Match 14.9%; Score 73.5; DB 1; Length 440;
Best Local Similarity 30.0%; Pred. No. 10;
Matches 18; Conservative 15; Mismatches 22; Indels 5; Gaps 1;

Qy 38 SLOGQWRCAAGTAQAQAAVVRFOEAANKQKQELDEISTNIRQAG-----VOYSDRADEQQQ 92
Db 2 SNOGAFDGAANVSGSRVFEVYVGMQRQNETDTNPIRKSGSVFIRVPRNRMNQEMQR 61

RESULT 10
ID KINH_DROME STANDARD; PRT; 975 AA.
AC P17210;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE KINESIN HEAVY CHAIN.
GN KHC OR KIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89168428; PubMed=25223352;
RA Yang J.T., Laymon R.A., Goldstein L.S.B.;
RT "A three-domain structure of kinesin heavy chain revealed by DNA
RL sequence and microtubule binding analyses.";
RL Cell 56:879-889(1989).
RN [2]
RP MUTAGENESIS.
RX MEDLINE=93030741; PubMed=1384131;
RA Gho M., McDonald K., Ganetzky B., Saxton W.M.;
RT "Effects of kinesin mutations on neuronal functions.";
RL Science 258:313-316(1992).
CC -!- FUNCTION: KINESIN IS A MICROTUBULE-ASSOCIATED FORCE-PRODUCING
CC PROTEIN THAT MAY PLAY A ROLE IN ORGANELLE TRANSPORT.
CC -!- SUBUNIT: OLIGOMER COMPOSED OF TWO HEAVY CHAINS AND TWO LIGHT
CC CHAINS.
CC -!- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A LARGE GLOBULAR N-
CC TERMINAL DOMAIN WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY OF

```

```

CC KINESIN (IT HYDROLYZES ATP AND BINDS MICROTUBULE), A CENTRAL
CC ALPHA-HELICAL COILED COIL DOMAIN THAT MEDIATES THE HEAVY CHAIN
CC DIMERIZATION; AND A SMALL GLOBULAR C-TERMINAL DOMAIN WHICH
CC INTERACTS WITH OTHER PROTEINS (SUCH AS THE KINESIN LIGHT CHAINS),
CC VESICLES AND MEMBRANOUS ORGANELLES.
CC -!- MISCELLANEOUS: MUTANT FLIES DISPLAY IMPAIRED ACTION POTENTIAL
CC PROPAGATION AND NEUROTRANSMITTER RELEASE AT NEUROMUSCULAR
CC JUNCTIONS, BUT ARE STILL CAPABLE OF TRANSPORTING CERTAIN
CC MEMBRANES, INCLUDING SYNAPTIC VESICLES, TO THE NERVE TERMINAL.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; M24441; AAA28652.1; -.
CC DR PIR; A31497; A31497.
CC DR HSPSP; P56536; 2KIN.
CC DR FlyBase; FBgn0001308; Khc.
CC DR InterPro; IPR001752; -.
CC DR Pfam; PF00225; kinesin; 1.
CC DR PRINTS; PR00380; KINESINHEAVY.
CC DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
CC KW Motor protein; Microtubules; ATP-binding; Coiled coil.
CC FT DOMAIN 1 334 MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).
CC FT DOMAIN 335 931 COILED COIL.
CC FT DOMAIN 932 975 GLOBULAR.
CC FT DOMAIN 180 321 MICROTUBULE-BINDING.
CC FT NP_BIND 92 99 ATP (BY SIMILARITY).
CC FT SEQUENCE 975 AA; 110429 MW; 9966CBC35BA74FD6 CRC64;

Query Match 14.9%; Score 73.5; DB 1; Length 975;
Best Local Similarity 24.8%; Pred. No. 24;
Matches 31; Conservative 19; Mismatches 32; Indels 43; Gaps 5;

Qy 1 MAEMKTDATILGQEGAGN-----FERISGLKTDQDOVESTAGSLQGWGRGA 46
Db 618 ISKMKTEAKNIAORCSNMETQADSNKKISEYKDLGELLISQHEARKMSQESMR-- 675
Qy 47 AGTAAQAQAAVVRFOEAANKQKQELDEISTNIRQAGVQYSRAD-----EEQQQA-- 93
Db 676 -----EAENK-KRTLEEQIDSLREECALKAAEHVSAVNAEKKRAEELRSM 721

Qy 94 LSSQM 98
Db 722 FDSQM 726

RESULT 11
ID PHP_DROME STANDARD; PRT; 1589 AA.
AC P39769;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE POLYHOMEOTIC-PROXIMAL CHROMATIN PROTEIN.
GN PH-P.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-Imaginal disks;
RX MEDLINE=92146957; PubMed=13466609;

```

RA Decamilis M., Cheng N.S., Pierre D., Brock H.W.;  
RT "The polyhomeotic gene of Drosophila encodes a chromatin protein that  
RT shares polytene chromosome-binding sites with Polycomb";  
RL Genes Dev. 6:223-232(1992).  
RN [2]  
RP SEQUENCE OF 199-1584 FROM N.A.  
RX MEDLINE=9203031; PubMed=1937015;  
RA Deatrick J., Daly M., Randsholt N.B., Brock H.W.;  
RT "The complex genetic locus polyhomeotic in Drosophila melanogaster  
RT potentially encodes two homologous zinc-finger proteins";  
RL Gene 105:185-195(1991).  
CC -!- FUNCTION: BINDS TO POLYTENE CHROMOSOMES. SEEMS TO INTERACT WITH  
CC PC. MAY INTERACT WITH PROTEINS ALREADY BOUND TO PROMOTER  
CC COMPLEXES AND MAY BE A NEGATIVE REGULATOR OF HOMEOTIC AND  
CC SEGMENTATION GENES. PLAYS A ROLE IN REGULATING THE EXPRESSION OF  
CC OTHER PAIR-RULE GENES SUCH AS EVE, FTZ, AND H.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- TISSUE SPECIFICITY: SALIVARY GLANDS.  
CC -!- SIMILARITY: TO MOUSE EARLY DEVELOPMENT REGULATOR PROTEIN RAR-28.  
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; X63672; CAA45211.1; -  
DR EMBL; M64750; -; NOT\_ANNOTATED\_CDS.  
DR PIR: S23632; S23632.  
DR Flybase: FBgn0004861; ph-p.  
DR InterPro: IPR001660; -  
DR Pfam: PF00536; SAM; 1.  
DR Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.  
FT DOMAIN 74 80 POLY-GLN.  
FT DOMAIN 411 450 GLN-RICH.  
FT DOMAIN 494 520 GLN-RICH.  
FT DOMAIN 619 650 GLN-RICH.  
FT DOMAIN 775 960 GLN-RICH.  
FT DOMAIN 1233 1290 SER/THR-RICH.  
FT ZN\_FING 1365 1387 C4-TYPE.  
FT CONFLICT 254 254 MISSING (IN REF. 2).  
FT CONFLICT 1415 1415 D -> A (IN REF. 2).  
SQ SEQUENCE 1589 AA; 167297 MW; A6DF0CF9106E1891 CRC64;

Query Match 14.8%; Score 73; DB 1; Length 1589;  
Best Local Similarity 27.3%; Pred. No. 45;  
Matches 35; Conservative 20; Mismatches 35; Indels 38; Gaps 7;  
QY 1 MAEKTKTAAFLQGEANGFERISGDL-----KTQIDQVESTAG-----SL 39  
DB 703 LATLKE---IGVAGQ-NKVVGHLLTVQQQQATNLQVVAAGNKMVMSTGTPTIL 758  
QY 40 QGWRGAAGTAA-----QAAVVRQEAANKQ---KQELDEISTNRAGVQYRADEE 89  
DB 759 QNGQTHAATAAGVDKQQQLQFLFKQQLIQQQQMLQQQIAAQMQQQQAAVQ---AQQQ 815  
QY 90 QQQALSSQ 97  
DB 816 QQQQVSSQ 823  
RESULT 12  
PYS1\_MASLA  
ID PYS1\_MASLA STANDARD; PRT; 80 AA.  
AC P11396;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)

DE PHYCOBILISOME 8.9 KDA LINKER POLYPEPTIDE, PHYCOCYANIN-ASSOCIATED, ROD  
DE (ROD CAPPING LINKER PROTEIN).  
GN CPCD.  
OS Mastigocladus laminosus (Fischerella sp.).  
OC Bacteria; Cyanobacteria; Stigonematales; Fischerella.  
OX NCBI\_TaxID=1191;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kufer W., Hoegner A., Eberlein M., Mayer K., Buchner A.,  
RA Gottschalk L.;  
RL Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE.  
RX MEDLINE=86050914; PubMed=3933528;  
RA Fueglistaller P., Suter F., Zuber H.;  
RT "Linker polypeptides of the phycobilisome from the cyanobacterium  
RT Mastigocladus laminosus: amino-acid sequences and relationships";  
RL Biol. Chem. Hoppe-Seyler 366:993-1001(1985).  
CC -!- SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; M75599; AAC64652.1; -  
DR PIR: A24691; A24691.  
DR InterPro: IPR001685; -  
DR Pfam: PF01383; CpcD; 1.  
KW Phycobilisome; Photosynthesis.  
SQ SEQUENCE 80 AA; 9109 MW; 5CEAD36240A465F9 CRC64;  
Query Match 14.7%; Score 72.5; DB 1; Length 80;  
Best Local Similarity 29.3%; Pred. No. 2.1;  
Matches 22; Conservative 10; Mismatches 22; Indels 21; Gaps 2;  
QY 23 GDLKTDQIDQVESTAGSLQGWGGAAGTAAQAAVVRQEAANKQKQELDEISTNRAG-- 80  
DB 3 GQTTGLGDSVSSA-----SRVFEVGVGMQRQENDKNKYNIIRSGV 46  
QY 81 ---VOYSRADEEQOQ 92  
DB 47 YITVPYNRMSEMQR 61  
RESULT 13  
FLA2\_CAMJE  
ID FLA2\_CAMJE STANDARD; PRT; 575 AA.  
AC P22251;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE FLAGELLIN A.  
GN FLAA.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=81116;  
RX MEDLINE=91009243; PubMed=2211662;  
RA Nuijten P.J., van Asten F.J., Gastra W., van der Zeijst B.A.;  
RT "Structural and functional analysis of two Campylobacter jejuni  
RT flagellin genes";  
RL J. Biol. Chem. 265:17798-17804(1990).  
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.  
CC -!- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.

```

Query Match      14.7%; Score 72.5; DB 1; Length 1679;
Best Local Similarity 22.5%; Pred. No. 53;
Matches 20; Conservative 18; Mismatches 42; Indels 9; Gaps

QY 9 ATUGAEGNFERISGDLKTIDQVESTAGSLQGWGRGAAGTAQAQAVVR---FQEAANKQ 65
DB 778 AKIKLENNLERLSQDQKIQIEISIRCKDSQLKWAQNTIDDTMKMKSLLTSLNSKE 837
      | : : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
QY 66 -----KQELDEISTNIRQAGVOYSGRADE 88
DB 838 TTTEKLSSEINLQDKELRKTGKFOYKFLDQ 866
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 15
ID K2M3_SHEEP STANDARD; PRT; 502 AA.
AC P25691;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE KERATIN, TYPE II; MICROFIBRILLAR, COMPONENT 5.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RC STRAIN=MERINO; TISSUE=Wool;
RX MEDLINE=921171866; PubMed=1371668;
RA Sparrow L.G., Robinson C.P., Caine J., McMahon D.T.W., Strike P.M.;
RT "Type II intermediate-filament proteins from wool. The amino acid
RL sequence of component 5 and comparison with component 7c.";
RT Biochem. J. 282:291-297(1992).
CC -!- FUNCTION: WOOL MICROFIBRILLAR KERATIN.
CC -!- TISSUE SPECIFICITY: HARD KERATIN WOOL.
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFIBRILLAR KERATIN, I AND II (40-55 AND 56-70 KILODALTONS,
CC RESPECTIVELY). AT LEAST ONE TYPE OF CYTOSKELETAL KERATIN IS
CC PRESENT IN ALL VERTEBRATE EPITHELIAL CELLS.
CC -!- MISCELLANEOUS: THE LOW-SULFUR PROTEINS, DERIVED FROM THE
CC MICROFIBRILLAR FRACTION OF WOOL EXTRACTS, ARE COMPOSED OF TWO
CC FAMILIES, EACH CONSISTING OF 4 HOMOLOGOUS SUBUNITS: THE TYPE I
CC COMPONENTS (8C-1, 8C-2, 8A & 8B) AND THE TYPE II COMPONENTS (5,
CC 7A, 7B, AND 7C).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC InterPro: IPR001664; -.
DR InterPro: IPR003054; -.
DR Pfam: PF00038; filament; 1.
DR PROSITE: PRO1276; TYPE2KERATIN.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin.
FT MOD_RES 1 1 BLOCKED.
FT DOMAIN 1 122 HEAD.
FT DOMAIN 1 122 HEAD.
FT DOMAIN 123 429 ROD.
FT DOMAIN 430 502 TAIL.
FT DOMAIN 123 157 COIL 1A.
FT DOMAIN 158 167 COIL 1A.
FT DOMAIN 168 268 COIL 1B.
FT DOMAIN 269 285 COIL 1B.
FT DOMAIN 286 429 COIL 2.
FT UNSURE 1 2
FT UNSURE SC -> CS.
SQ SEQUENCE 502 AA; 55255 MW; 8734C8230550CE68 CRC64;

Query Match      14.6%; Score 72; DB 1; Length 502;
Best Local Similarity 23.0%; Pred. No. 16;
Matches 20; Conservative 19; Mismatches 36; Indels 12; Gaps

QY 17 NFERISGDLKTIDQVESTAGSLQGWGRGAAGTAQAQAVVRFQEAANKQKQELDE----- 71
DB 286 NMDCVVAETKAOVDITASRSRAEASWYSKCEEMKATVIRGETLRRTKKEINELNRVI 345
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Thu Jul 5 13:58:57 2001

us-09-462-480-5.rsp

Page 9

QY 72 --ISTNIRQAGVQYSR-----ADEEQQ 91  
Db 346 QRLTAIEIENAKCORTKLEAAVAEEQQ 372

Search completed: July 5, 2001, 11:51:41  
Job time: 500 sec

THIS PAGE BLANK (USPTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 11:42:57 ; Search time 123.78 seconds  
(without alignments)  
106.887 Million cell updates/sec

Title: US-09-462-480-5  
Perfect score: 492  
Sequence: 1 MAEMKTDAAITLQEGAGNFER.....VOYSGRADEQOALLSQMGF 100

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPREMBL\_16.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_unclassified.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_virus.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	486	98.8	100	2	069739
2	187	38.0	100	2	033084
3	92.5	18.8	1186	2	092802
4	91.5	18.6	410	2	093367
5	90	18.3	750	3	042657
6	89	18.1	103	2	005440
7	87.5	17.8	97	2	053692
8	84.5	17.2	97	2	053267
9	82.5	16.8	188	5	077248
10	81	16.5	528	5	026589
11	81	16.5	1940	5	002456
12	79.5	16.2	163	2	099446
13	79.5	16.2	1999	11	063731
14	79	16.1	245	2	069467
15	79	16.1	508	6	028582
16	77	15.7	178	5	002415
17	77	15.7	507	4	P78386
18	77	15.7	507	4	Q9NSB1
19	77	15.7	561	2	Q91016

20	77	15.7	1326	2	Q9L2C3
21	76.5	15.5	528	2	Q9RSJ1
22	76.5	15.5	707	2	Q9FSN1
23	76.5	15.5	842	5	Q9VEH0
24	76.5	15.5	1278	9	Q9XJA8
25	76	15.4	331	2	Q9X7H6
26	76	15.4	1530	4	O43241
27	76	15.4	2756	10	Q9LJ60
28	75.5	15.3	1961	11	O62812
29	75	15.2	507	11	Q922T6
30	75	15.2	671	13	Q9YHD4
31	74.5	15.1	242	1	Q9HHC2
32	74.5	15.1	387	2	Q54842
33	74.5	15.1	548	2	Q9EWY9
34	74.5	15.1	576	2	Q9R950
35	74.5	15.1	880	2	Q9I245
36	74.5	15.1	929	3	P78718
37	74.5	15.1	1313	10	Q9XIP6
38	74.5	15.1	1956	5	Q20641
39	74	15.0	493	4	Q9NSB3
40	74	15.0	756	2	Q9RYG4
41	74	15.0	1447	11	Q9QYT3
42	74	15.0	1487	11	Q9QYT2
43	74	15.0	3663	5	Q9TW65
44	74	15.0	3674	5	Q9TYG9
45	73.5	14.9	388	2	Q53474

## ALIGNMENTS

RESULT	1
069739	
ID	069739 PRELIMINARY; PRT; 100 AA.
AC	069739;
DT	01-AUG-1998 (TREMELREL. 07, Created)
DT	01-AUG-1998 (TREMELREL. 07, Last sequence update)
DT	01-JUN-2000 (TREMELREL. 14, Last annotation update)
DE	HYPOTHETICAL 10.8 KDA PROTEIN.
GN	LHP OR RV3874 OR MTV027.09.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1773;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE=98295987; PubMed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA	Taylor K., Whitehead S., Barrell B.G.;
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RL	complete genome sequence."
RN	Nature 393:537-544(1998).
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RA	Berthet F.X., Birk Rasmussen P., Andersen P., Gicquel B.;
RT	"Promoter analysis of the M. tuberculosis orfC gene encoding the
RT	early secreted antigenic target 6 kDa (ESAT-6)."
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AL022120; CAA17966.1;
DR	EMBL; AF004671; AAC83445.1;
DR	TubercuList; RV3874;
KW	Hypothetical protein.
SQ	SEQUENCE 100 AA; 10794 MW; 285F4FC9F55D194 CRC64;

Q9L2C3	streptomycin
Q9RSJ1	deinococcus
Q9FSN1	rhizobium m
Q9VEH0	drosophila
Q9XJA8	streptococ
Q9X7H6	paracoccus
O43241	homo sapien
Q9LJ60	arabidopsis
O62812	rattus norv
Q922T6	mus musculu
Q9YHD4	rana catesb
Q9HHC2	halobacteri
Q54842	streptococ
Q9EWY9	streptococ
Q9R950	campylobact
Q9I245	pseudomonas
P78718	nectria hae
Q9XIP6	arabidopsis
Q20641	caenorhabdi
Q9NSB3	homo sapien
Q9RYG4	deinococcus
Q9QYT3	mus musculu
Q9QYT2	mus musculu
Q9TW65	caenorhabdi
Q9TYG9	caenorhabdi
Q53474	streptococ

Query Match 98.8%; Score 486; DB 2; Length 100;  
Best Local Similarity 99.0%; Pred. No. 4e-38;  
Matches 99; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAEMKTDAAITGQAGNFERISGDLKTDQVSTAGSLQGWGAGTAQAQAAVVRFOE 60  
DB 1 MAEMKTDAAITGQAGNFERISGDLKTDQVSTAGSLQGWGAGTAQAQAAVVRFOE 60  
QY 61 AANKQKQELDEISNIRAGVOYSGRADEEQOQALSSQMGF 100  
DB 61 AANKQKQELDEISNIRAGVOYSGRADEEQOQALSSQMGF 100

RESULT 2  
O33084 PRELIMINARY; PRT; 100 AA.  
AC O33084; 1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-JAN-1998 (Tremblrel. 08, Last annotation update)  
DE SIMILARITY TO SMALL.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN SEQUENCE FROM N.A.  
RP Eglmeier K., Garnier T., De Rossi E., Fsihi H., Cole S.T.;  
RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93188700; PubMed=8446027;  
RA Eglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;  
RT "Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae."  
RL MOL. Microbiol. 7:197-206(1993).  
DR EMBL; Y14967; CAA75210.1; -  
SQ SEQUENCE 100 AA; 10964 MW; 460EE12F876EC383 CRC64;

Query Match 38.08%; Score 187; DB 2; Length 100;  
Best Local Similarity 40.08%; Pred. No. 2e-10;  
Matches 40; Conservative 20; Mismatches 40; Indels 0; Gaps 0;

QY 1 MAEMKTDAAITGQAGNFERISGDLKTDQVSTAGSLQGWGAGTAQAQAAVVRFOE 60  
DB 1 MAEMTEAAILTQQAQFDQIASGLSQERNFVDSIQSFQNTWEGQAASALGALGRFDE 60  
QY 61 AANKQKQELDEISNIRAGVOYSGRADEEQOQALSSQMGF 100  
DB 61 AMQDQIRQLESIVDKLNRSGGNYTKTDDEANQLLSKRMNF 100

RESULT 3  
O92BQ2 PRELIMINARY; PRT; 1186 AA.  
AC O92BQ2;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE PUTATIVE CHROMOSOME ASSOCIATED PROTEIN.  
GN SC/Al.21.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN SEQUENCE FROM N.A.  
RP STRAIN-A3(2);  
RC Murphy L., Harris D.;  
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RL [2]  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.;  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
RL MOL. Microbiol. 21:77-96(1996).  
DR EMBL; AL034447; CAA22420.1; -  
DR InterPro; IPR001066; -  
DR InterPro; IPR001687; -  
DR InterPro; IPR003395; -  
DR InterPro; IPR003405; -  
DR InterPro; IPR003439; -  
DR Pfam; PF02463; SMC\_N; 1.  
DR Pfam; PF02483; SMC\_C; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
SQ SEQUENCE 1186 AA; 128722 MW; CB11027815373E99 CRC64;

Query Match 18.8%; Score 92.5; DB 2; Length 1186;  
Best Local Similarity 32.3%; Pred. No. 1.8;  
Matches 32; Conservative 16; Mismatches 44; Indels 7; Gaps 3;  
QY 2 AEMKTDAAITGQAGNFERISGDLKTDQVSTAGSLQGWGAGTAQAQAAVVRFOE 59  
DB 686 AGRRECAALVEELGERRRAADREKSSVAQ---QLGSLAGQARGAAGEAERSAAAEAAQ 742  
QY 60 EAANKQKQELDEISNIRAGVOYSGRADEEQOQALSSQMGF 98  
DB 743 EALDKALTEVEELAE--RLAVAEEMPVEEEDPTAARDRL 779

RESULT 4  
O9X367 PRELIMINARY; PRT; 410 AA.  
AC O9X367;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
DE PX01-98.  
OS Bacillus anthracis.  
OG Plasmid virulence plasmid PX01.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1392;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-STERNE;  
RA Okinaka R.T., Cloud K., Hamton O., Hoffmaster A., Hill K.K., Keim P.,  
RA Koehler T., Lamke G., Kumano S., Mahillon J., Manter D., Martinez Y.,  
RA Ricke D.O., Svensson R., Jackson P.J.;  
RT "The sequence and organization of pX01, the large Bacillus anthracis  
plasmid harboring the Anthrax toxin genes."  
RL J. Bacteriol. 0:0-0(1999).  
DR EMBL; AF065404; AAD32402.1; -  
DR InterPro; IPR001680; -  
DR PROSITE; PS00678; WD\_REPEATS; UNKNOWN\_1.  
KW Plasmid.  
SQ SEQUENCE 410 AA; 45798 MW; F10BC04607575CTD CRC64;

Query Match 18.8%; Score 91.5; DB 2; Length 410;  
Best Local Similarity 28.0%; Pred. No. 0.69;  
Matches 26; Conservative 17; Mismatches 29; Indels 21; Gaps 3;  
QY 3 EMKTDAAITGQAGNFERISGDLKTDQVSTAGSLQGWGAGTAQAQAAVVRFOEAA 62  
DB 11 OLEQAATVNRSSLEYIHODLYSQTEYIAS-----QWSGASSD-----RFTQMF 56



DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE PE-FAMILY PROTEIN.  
 GN RV3020C OR MV012.34C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 DR EMBL; AL021287; CAA16105.1; -  
 DR TuberculList; RV3020C; -  
 SQ SEQUENCE 97 AA; 9842 MW; BA9CB3180EC17F2 CRC64;

Query Match 17.28; Score 84.5; DB 2; Length 97;  
 Best Local Similarity 34.99; Pred. No. 0.62; Indels 3; Gaps 1;  
 Matches 29; Conservative 7; Mismatches 44;  
 QY 18 PERISGDLKTDIDQVESTAGSLQGWGAGTAAGTAAGAAVVFQEAANKQKQELDEISTNIR 77  
 Db 18 FAAGAGLMRHHTIGAEQAQMSAQAFQGESAAAFQGAHAREFAAAKNTLLDIAQANLG 77  
 QY 78 QAGVOYSRADEEQQAALSSQMGF 100  
 Db 78 EAAGTYVAAD---AAAASSYTG 97

RESULT 9  
 O77248 ID O77248 PRELIMINARY; PRT; 188 AA.  
 AC O77248;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE APOLIPOPHORIN-III PRECURSOR.  
 GN APOLP-III.  
 OS Spodoptera litura.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.  
 OX NCBI\_TaxID=69820;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kim E., Kim S.H., Choi C.S., Park Y.I., Kim H.R.;  
 RA "Cloning and Expression of Apolipophorin-III from the Common Cutworm,  
 RT Spodoptera litura."  
 RL Arch. Insect Biochem. Physiol. 0:0-0(1998).  
 DR EMBL; AF094582; AAC63377.1; -  
 SQ SEQUENCE 188 AA; 20649 MW; 7A8AE366DF8E432C CRC64;

Query Match 16.98; Score 82.5; DB 5; Length 188;  
 Best Local Similarity 26.78; Pred. No. 2;  
 Matches 27; Conservative 20; Mismatches 45; Indels 9; Gaps 2;  
 QY 6 TDNAKAKTALEQARQNLKTAEDLRKSHDPVQRQAGELRTKLQAAVONTAQEVQKLAKE 147  
 Db 88 TDNAKAKTALEQARQNLKTAEDLRKSHDPVQRQAGELRTKLQAAVONTAQEVQKLAKE 147

QY 61 AANKQKQELDEISTNIRQAGVOYSRADEEQ---QALSSQ 97  
 Db 148 VASNVETNEKLPKLEAVENFSKHVEEVQKQKVEAASKQ 188  
 RESULT 10  
 Q26589 ID Q26589 PRELIMINARY; PRT; 528 AA.  
 AC Q26589;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE MYOSIN II HEAVY CHAIN (FRAGMENT).  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;  
 OC Trematoda; Digenea; Strigeida; Schistosomatidae; Schistosomatidae;  
 OC Schistosoma.  
 OX NCBI\_TaxID=6183;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PUERTO RICAN;  
 RA Amory L.M.;  
 RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PUERTO RICAN;  
 RX MEDLINE=93056536; PubMed=1431131;  
 RA Soisson L.M., Masterson C.P., Tom T.D., McNally M.T., Lowell G.H.,  
 RA Strand M.;  
 RT "Induction of protective immunity in mice using a 62-kDa recombinant  
 RT fragment of a Schistosoma mansoni surface antigen."  
 RL J. Immunol. 149:3612-3620(1992).  
 DR EMBL; X65591; CAA46548.1; -  
 DR HSP; P08799; IIMD.  
 DR InterPro; IPR000048; -  
 DR InterPro; IPR000533; -  
 DR InterPro; IPR001609; -  
 DR InterPro; IPR001637; -  
 DR InterPro; IPR002928; -  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR PRINTS; PR00194; TROPOMYOSIN.  
 DR PROSITE; PS00182; GLNA\_ADENYLATION; UNKNOWN1.  
 DR SMART; SM00015; IQ; 1.  
 KW Myosin.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 528 AA; 61622 MW; AF075D13EB249B4C CRC64;

Query Match 16.5%; Score 81; DB 5; Length 528;  
 Best Local Similarity 21.9%; Pred. No. 8.5;  
 Matches 21; Conservative 24; Mismatches 45; Indels 6; Gaps 2;  
 QY 3 EMKTDAAITLQGEAGNFERISGLKTDIDQVESTAGSLQGWGAGTAAGAAVVFQEA 62  
 Db 337 KLEGDLKATQETVDDLBRVKRDLEQLRRKAEIGLGGKFEQGLVAQ-----LQK 391  
 QY 63 NKQKQELDEISTNIRQAGVOYSRADEEQQAALSSQ 98  
 Db 392 KELQTRIQELEDLEAERAARSRAKRSRQ-LESEL 426

RESULT 11  
 Q02456 ID Q02456 PRELIMINARY; PRT; 1940 AA.  
 AC Q02456;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE MYOSIN HEAVY CHAIN.  
 GN MYH.

```

Q9RY46;
01-MAY-2000 (TrEMBLrel. 13, Created)
01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 17.8 KDA PROTEIN.
DR0105.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
ON NCBI_TaxID=1299;
RX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Otterbeck T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).
DR EMBL: AEO01873; AAF09701.1; -.
DR TIGR: DR0105; -.
KW Hypothetical protein.
SQ SEQUENCE 163 AA; 17830 MW; A0AC12CE20534D9C CRC64;

Query Match 16.2%; Score 79.5; DB 2; Length 163;
Best Local Similarity 22.7%; Pred. No. 3.2;
Matches 22; Conservative 16; Mismatches 46; Indels 13; Gaps

QY 11 LGQFAGN-----FETISGDLKQTIDQVESTAGSLQGWGCAAGTAAQAAVVR 57
Db 1 MGDEAGNAQDRAEAAARAESHGADYRRMDSANDDLLSGLGQAKHKTKDQAAEKVRAD 60
QY 58 FQEAANKQKQELDEISTNIRQAGVOYSRADDEEQQAL 94
Db 61 TQEAQNAKRAQDVNRANVHESAQDFRAGAEQAQTL 97

RESULT 13
Q63731
ID Q63731 PRELIMINARY; PRT; 1999 AA.
IC Q63731;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE NEURONAL MYOSIN HEAVY CHAIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
ON NCBI_TaxID=10116;
RX [1]
RN SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=92235856; PubMed=1569576;
RA Sun W., Chantler P.D.;
RT "Cloning of the cDNA encoding a neuronal myosin heavy chain from
RT mammalian brain and its differential expression within the central
RT nervous system[see comments].";
RT J. Mol. Biol. 224:1185-1193(1992).
DR HSSP: P10587; IBB1.
DR InterPro: IPR000048; -.
DR InterPro: IPR001609; -.
DR InterPro: IPR002017; -.
DR InterPro: IPR002928; -.
DR Pfam: PF000063; myosin_head; 1.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF01576; Myosin_tail; 2.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; -. 1.

```

DR SMART; SM00015; IQ; 1.  
KW Myosin.  
SQ SEQUENCE 1999 AA; 231456 MW; 090C181D55727B41 CRC64;

Query Match 16.2%; Score 79.5; DB 11; Length 1999;  
Best Local Similarity 23.1%; Pred. No. 51;  
Matches 31; Conservative 26; Mismatches 28; Indels 49; Gaps 5;

QY 9 ATLGQAGN-----PERISGLKTDQVSTAGSLQGWGGAAGTAQAQAVVRF 58  
DB 1764 ADLNLGRHAQKNENARQQLERQNKELKVLQEME---GVKSYK-ASITALEAKTAQL 1819  
QY 59 QE-----AAKQ-----KOELDEISTNIRAGVQY 83  
DB 1820 EEQLDNETKERQAACKQVRTEKKLKVLLQVDDDERNAEQYKDAQKASTRLKQLKRL 1879  
QY 84 SRADEEQQALSSQ 97  
DB 1880 EEAEQAQRANASR 1893

## RESULT 14

O69467 PRELIMINARY; PRT; 245 AA.  
AC O69467;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
DE HPOTHEICAL 27.1 KDA PROTEIN.  
GN MLCB1243.13.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID-1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hamlin N., Churcher C.M.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-93188700; PubMed-8446027;  
RA Eiglmeyer K., Honore N., Woods S.A., Caudron B., Cole S.T.;  
RT "Use of an ordered cosmid library to deduce the genomic organization  
of Mycobacterium leprae";  
RL Mol. Microbiol. 7:197-206(1993).  
DR EMBL; AL023635; CAA19194.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 245 AA; 27087 MW; C984D9A5FA49697A CRC64;

Query Match 16.1%; Score 79; DB 2; Length 245;  
Best Local Similarity 25.7%; Pred. No. 5.6;  
Matches 28; Conservative 20; Mismatches 45; Indels 16; Gaps 4;

QY 2 AEMKTDATLGQEA-----GNPERISGLKTDQVSTAGS-----LQGWGGAAGTAA 51  
DB 70 ADSMVSSATTESLLSHARAEADRLISDAKSQVDRMASEARQHSERMGLGDAREESIRIA 129  
QY 52 QAQAVVRFQEAANKOKQELDEISTNIRAGVQYSRADDE---QQQALSSQ 97  
DB 130 TVAKREYEASLNRAQSECDRL---IENGNISYEKAIQEGIKQORLVSQ 175

## RESULT 15

Q28582 PRELIMINARY; PRT; 508 AA.  
ID Q28582  
AC Q28582;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE HAIR TYPE II KERATIN INTERMEDIATE FILAMENT PROTEIN.  
GN KII-9.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID-9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-92274852; PubMed-1375545;  
RA Powell B., Crocker L., Rogers G.;  
RT "Hair follicle differentiation: expression, structure and evolutionary  
conservation of the hair type II keratin intermediate filament gene  
family";  
RL Development 114:417-433(1992).  
CC -I- SIMILARITY: TO THE INTERMEDIATE FILAMENT FAMILY.  
DR EMBL; X62509; CAA44368.1; -;  
DR InterPro: IPR001664; -;  
DR InterPro: IPR003054; -;  
DR Pfam: PF00038; filament; 1.  
DR PRINTS; PR01276; TYPE2KERATIN.  
DR PROSITE; PS00226; IF; 1.  
KW Coiled coil; Heptad repeat pattern; Intermediate filament; Keratin.  
SQ SEQUENCE 508 AA; 55245 MW; 3E728AA9512F96F7 CRC64;

Query Match 16.1%; Score 79; DB 6; Length 508;  
Best Local Similarity 22.9%; Pred. No. 13;  
Matches 19; Conservative 23; Mismatches 39; Indels 2; Gaps 1;

QY 17 NFERISGLKTDQVSTAGSLQGWGGAAGTAQAQAVVRFQEAANKOKQELDEISTNI 76  
DB 271 NMNIVAEIKAQYDDIASRSRAEASWYRSKCEIKATVIRHGETLRTKEEINELNRVI 330  
QY 77 RQ--AGVQYSRADEEQQALSSQ 97  
DB 331 QRLTAEEVENAKCQNSKLEAAVTQ 353

Search completed: July 5, 2001, 11:50:46  
Job time: 469 sec

Thu Jul 5 13:58:58 2001

us-09-462-480-5.rspt

Page 7

**THIS PAGE BLANK (USPTO)**